

THIS WEEK

EDITORIALS

FISHY The continuing tale of how to measure the ones that got away **p.282**

WORLD VIEW To save bees or not to save bees? **p.283**



BRIGHT START Light nights accelerate bird reproduction capacity **p.284**

Vital statistics

That robust data are not collected on births, deaths and causes of death is a scandal. A new drive and greater investment are needed to grow the field of health metrics.

Many readers of *Nature* will take it for granted that they have a birth certificate, and that when they die, their death, and its cause, will be officially recorded, as will their health problems in the intervening years. When aggregated, such data allow researchers to estimate disease burdens and risks to help shape public-health policies and investment in everything from high blood pressure to infectious diseases — and to monitor the impact of disease control efforts.

Yet more than 100 countries, and not just the poorest, lack even basic birth and death registration systems. Furthermore, only 34 nations — covering just 15% of the global population — generate decent cause-of-death data, and even some of those data are unreliable because doctors have not correctly ascribed the cause of death.

There is a shocking lack of national and international political will to invest in the basic statistical systems needed to track this most fundamental information. Bodies such as the World Health Organization (WHO) continue to push out charts of global trends. These are handy for advocacy purposes, but the underlying data are often scarce and poor.

Initiatives such as the Global Burden of Disease study — published in *The Lancet* last December by an international consortium led by the Institute for Health Metrics and Evaluation in Seattle, Washington — have helped. They have sucked up what data are available from demographic health surveys, papers and other sources, and brought unparalleled scientific expertise and advanced modelling to bear on extracting meaning from the sparse and heterogeneous data — and filled in gaps where no data exist at all (see *Nature* 492, 311–312; 2012). But even the researchers involved are the first to admit that this situation is far from ideal, and that what is really needed is more and better raw data.

The issue of how to improve global health estimates was the subject of a two-day meeting convened in Geneva, Switzerland, last week by the WHO. Many people thought the meeting was constructive, although the consensus recommendations that emerged — for the WHO and academics to collaborate more closely; increased investment in registration systems and training; and better sharing of data and methods — will need to be accompanied by consolidated political commitment to gathering health metrics.

Although their intergovernmental nature and direct contact with ministries mean that the WHO and other United Nations (UN) agencies are essential players in getting better registration systems, they can also be part of the problem. Numerous agencies are involved in health metrics, but they are largely uncoordinated, overly bureaucratic and politicized and too oriented towards defending their turf. No one agency is responsible for promoting civil birth and death registration.

The latest disappointment is the Health Metrics Network (HMN), a WHO-hosted partnership of international organizations created in 2005 to boost civil registration health data with US\$50 million from the Bill & Melinda Gates Foundation. Despite a promising start,

observers say that there have been few accomplishments to show for the money, and the WHO dissolved the network last November.

That makes the goal to boost civil registration systems more necessary than ever. The new reality is that most of the expertise in health estimates is no longer within the UN; it is in academia. *Nature* has learnt that at the same time as the WHO meeting in Geneva, other leading scientists in the field were meeting with philanthropists in New York on how to replace the HMN with a new organization — one that would not be hosted within the WHO.

“The size of the field is incommensurate with the immense task at hand.”

A fundamental problem is that the size of the field is incommensurate with the immense task at hand, and that is further complicated by intense competition for limited funds. The community must work to better present its very justified case for

greater political attention and funding — and for a much needed injection of fresh blood and expertise, especially with a national focus.

Given the information technology of the twenty-first century, it is simply unacceptable that the relatively cheap and simple registration systems needed to gather data on births and causes of death on a continuous basis are absent across much of the planet. The development of such systems is largely the responsibility of individual nations, but greater political attention is needed at both the national and international levels to make it happen. A good place to start would be placing the seemingly mundane, yet crucial, issue of civil registration systems higher on the agenda of organizations such as the G20. ■

Eyes and ears

Two explosions last week demonstrated the importance of global monitoring.

On 15 February, the town of Chelyabinsk in the Russian Ural Mountains had an unexpected visitor. A meteor streaked high above the city, briefly blinded commuters and then shattered thousands of windows with a series of ear-splitting explosions. The event was recorded on mobile phones and car-dashboard cameras across the region, and YouTube soon filled with Hollywood-style disaster videos of the fireball, replete with some very colourful Russian commentary.

Local residents were not the only ones to record the blast. More than a dozen monitoring stations around the globe captured the ultra-low-frequency infrasound signal of the meteorite as it broke up in the atmosphere. The stations are part of a much larger network of sensors

that has been built to detect illicit nuclear testing: the system, it is hoped, will eventually underpin the Comprehensive Nuclear-Test-Ban Treaty (CTBT), a pact to halt work on nuclear weapons worldwide. Using data from this monitoring system, scientists in Canada and the United States were quickly able to establish that the rock that broke up over Russia was the largest to strike Earth in more than a century. They found that it exploded with the strength of a good-sized thermonuclear warhead, although, luckily, at an altitude high enough for the atmosphere to absorb most of the shock wave.

To understand the value of this monitoring network, imagine that the celestial visitor had arrived 30 years earlier — no time at all in the life of the Solar System. If there had been a sudden explosion over Chelyabinsk towards the end of the cold war, without an Internet or free press to circulate images, a very different picture could have emerged. The city is fewer than 100 kilometres from some of Russia's largest nuclear-weapon production and storage facilities: a surprise airburst would almost certainly have put the country's nuclear arsenal on hair-trigger alert. Shortly after the strike last week, right-wing lawmaker Vladimir Zhirinovskiy asserted: "Those aren't meteors falling, it's the Americans testing new weapons." His comments were greeted with bemusement by the Russian press; in another time, they might have triggered nuclear war.

Just a few days before the spectacular events over Russia, the CTBT network picked up a less visible but politically more significant incident. On 12 February, North Korea conducted its third nuclear-weapon test deep underground. On this occasion, the CTBT network's seismic sensors detected the blast, and located it to within a few kilometres of North Korea's previous nuclear tests. Independent analysis of the network's data showed the yield of the weapon to be several kilotonnes, much smaller than the explosion of the Russian meteor.

Unlike with the Russian event, there were few other ways to verify the North Korean explosion. The North Korean Central News Agency put out a statement announcing the test, but is not particularly reliable. US, Japanese and South Korean sensors all picked up the shock from the blast, but because they belong to sovereign nations, there was no guarantee that the data would be shared in a timely fashion — or believed by adversaries.

"A ban on nuclear testing could be enforced, if a further eight nations are willing to ratify it."

The *raison d'être* of the CTBT network is to catch tests such as the one conducted by North Korea. Its ability to do so shows that an international ban on nuclear testing could be enforced, if a further eight nations, including China, the United States, India and Pakistan, were willing to ratify it. The CTBT has been open for ratification since 1996, but unfortunately, in recent years, little progress has been

made towards its entry into force.

The meteor strike also shows that the constructed network has great value in its own right. It has done much non-nuclear-test work since it became active: tracking earthquakes, tsunamis and nuclear accidents.

Building and running this global sensor network isn't cheap. The CTBT organization in Vienna estimates that around US\$100 million a year goes on its 321 monitoring stations and 16 laboratories worldwide, along with a data centre and other support for the treaty. Those funds are contributed by the treaty organization's 183 member states, which are guaranteed timely access to the data collected by the network.

Many hundreds of scientists have begun using the CTBT data in the past few years, and many hundreds more are likely to sign up. As the events of the past week show, even without a test-ban treaty, the network makes the world a safer and more interesting place to live. ■

Net gains

Estimating the scale of the problem may allow us to arrest dangerous levels of overfishing.

The sea is a big place. Most fish are small. So it stands to reason that it is difficult to work out with any degree of accuracy just how many fish live in the sea. One way is to measure how many fish we pull out of it. But is that the best way? Or even an accurate way? In two Comment pieces this week, starting on page 303, fisheries scientists debate the issue. It is a crucial one. Worldwide, more than US\$200 billion of fish were caught or farmed in 2010. How long can that continue?

In one piece, Daniel Pauly argues that 'catch data' of the number of fish caught are a vital tool for assessing the health of fish stocks. In their counterpoint piece, Ray Hilborn and Trevor Branch warn that over-reliance on this measure misses important subtleties and can misleadingly distil the health of entire ecosystems down to a landed tonnage. This is far from an academic debate. If scientists cannot estimate fish numbers, and so the health of stocks, there is little hope that this resource can be exploited in a sustainable fashion.

Disagreements such as this can be problematic for policy-makers. They want a simple answer to the question of how much fish should be caught. But it is crucial that they happen, and happen openly. Fisheries science, and marine science generally, may never have been more important.

It is unquestionable that some fisheries have been horribly mismanaged, and some species driven to dangerously low levels. But equally, there are positive signs of change. There are examples of well-managed fisheries, and, more importantly, there now seems to be a political will to listen to scientists. In the past, quotas for fishing were frequently set much higher than recommended. Europe's rightly derided Common

Fisheries Policy (CFP) is a leading example of this. Tuna populations also show the dangers of repeatedly ignoring scientific advice.

Last year, the International Commission for the Conservation of Atlantic Tunas surprised some by sticking to scientific advice on how many of the valuable fish should be caught, despite suggestions that numbers are increasing. And European politicians are pushing for a reform of the CFP that may finally put science in the driving seat in setting catch limits. Schemes to tell consumers which fish they can eat with a clear conscience have never been more popular, and are also attracting increasing, healthy scrutiny (including in these pages; see J. Jacquet *et al.* *Nature* **467**, 28–29; 2010, and related Correspondence).

Marine conservation more broadly is also gathering pace. Huge marine reserves are being created around the world, although these are not without teething problems and whether they will ultimately boost fisheries is hotly debated. Billionaires vie to explore the depths, bringing with them slick technology, show-business élan and even more public attention. Last week saw the launch of the Global Ocean Commission, with senior political figures aiming to produce recommendations on how to preserve the ecosystems of the high seas outside national jurisdictions, to feed into United Nations discussions set for 2014.

One message from the Comment pieces this week is just how little reliable information we have about fisheries. Pauly admits that catch data are massively under-reported in many countries, and Hilborn and Branch cite the value of more-detailed scientific assessments of stock while acknowledging that these exist for only 40% of the total catch in the global database of the Food and Agriculture Organization of the UN.

Fisheries scientists unwilling to face this reality can take heart. We don't have the basic information to judge the health of many human stocks either (see page 281). Those who have the more difficult job of sifting the oceans must be brave enough to outline the uncertainties — such as those over catch data — even as they fight to reduce them. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Bees, lies and evidence-based policy

Misinformation forms an inevitable part of public debate, but scientists should always focus on informing the decision-makers, advises Lynn Dicks.

Saving bees is a fashionable cause. Bees are under pressure from disease and habitat loss, but another insidious threat has come to the fore recently. Concern in conservation and scientific circles over a group of agricultural insecticides has now reached the policy arena. Next week, an expert committee of the European Union (EU) will vote on a proposed two-year ban on some uses of clothianidin, thiamethoxam and imidacloprid. These are neonicotinoids, systemic insecticides carried inside plant tissues. Although they protect leaves and stems from attack by aphids and other pests, they have subtle toxic effects on bees, substantially reducing their foraging efficiency and ability to raise young.

Whatever the EU decision, this vote will not be the end of the story. The proposed ban will buy some time for scientists and policy-makers to understand more about how neonicotinoids affect bee populations. For despite what both sides of the argument say, the link between bee declines and neonicotinoids is far from clear. I gave evidence to a UK parliamentary inquiry on the issue late last year, and my experience offers a useful window on how science informs public debate and policy-making — and, in the case of the public debate, how it does not.

There is no doubt that the proposed restriction on the use of these neonicotinoids on nectar- and pollen-rich crops such as oilseed rape will reduce a potentially serious risk to bees. It seems a crucial step towards reversing or halting observed declines in bees and other flower-feeders. But that is not enough for some environmental campaigners, who have framed the problem as one of the very survival of an unspecified number of bee species. Two and a half million people have signed an online petition telling EU decision-makers: “If you act urgently with precaution now, we could save bees from extinction.”

The assertion that a ban on neonicotinoids in Europe will save bees from extinction is absurd. There are bee species around the world in genuine danger of extinction, such as the once-common rusty-patched bumblebee in the United States, which has vanished from 87% of its historic range since the early 1990s. Diseases, rather than pesticides, are suspected of driving that decline. And although there have been dramatic falls in the numbers of managed honey bee *Apis mellifera* colonies in some countries, it remains a widespread and common bee, not in imminent danger of extinction.

Well-meaning exaggeration is common. *The Guardian*, a pro-environment British newspaper, mangled my parliamentary evidence on moths and beetles to claim that three-quarters of all UK pollinator species, including bees, were in severe decline.

There are startling claims in favour of neonicotinoids too. One headline widely reported

in the UK farming press is that, without them, UK wheat yields could decline by up to 20%. This is a disingenuous interpretation of an industry-funded report, and the EU is not proposing to ban neonicotinoid use in wheat anyway, because wheat is not a crop attractive to bees.

As a scientist involved in this debate, I find this misinformation deeply frustrating. Yet I also see that lies and exaggeration on both sides are a necessary part of the democratic process to trigger rapid policy change. It is simply impossible to interest millions of members of the public, or the farming press, with carefully reasoned explanations. And politicians respond to public opinion much more readily than they respond to science.

There is a precedent here. The 1987 Montreal Protocol that banned chlorofluorocarbons to protect the ozone layer is commonly held up as a shining example of a rapid policy response to emerging science. Yet

it was agreed against a backdrop of wild stories of millions of extra cases of cancer and industry warnings that it would cost the US economy billions of dollars.

There is a risk, of course, that rapidly made, responsive policy changes will not turn out to be the most intelligent ones. We saw this in the European biofuels policy, which set a target of 10% renewable content in transport fuels by 2020, despite evidence at the time that this was not the best way to reduce greenhouse-gas emissions using renewable energy.

This risk means that communicating the science itself directly to appropriate decision-makers remains extremely important. Scientists must not be turned off by the rhetoric, but motivated by it.

We should engage with the debate throughout. It is important to get as near to the decision-makers as possible, providing clear and well-referenced information with an independent voice.

You can't switch off the lies and exaggeration. But don't worry about them. When I saw the exaggerated pollinator-decline claim attributed to me in *The Guardian* I did not seek to correct it, because the correct information, with references, will go into a forthcoming parliamentary-committee report. Unlike stories in the press, that report will definitely be read by officials who advise the politicians who, for the United Kingdom at least, make the final decision. And because of such reports, and a recent risk assessment from the European Food Safety Authority, we can be fairly sure that the decision on whether to restrict neonicotinoid use in Europe will not be made on the basis of avoiding 20% yield losses in crops, or saving the world's bees from extinction. ■

Lynn Dicks is a Natural Environment Research Council Knowledge Exchange Fellow at the University of Cambridge, UK. Read her blog at go.nature.com/gjondz
e-mail: lvd22@cam.ac.uk

SCIENTISTS MUST
NOT BE
TURNED OFF
BY THE RHETORIC, BUT
MOTIVATED
BY IT.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/sg7wap

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

MOLECULAR EVOLUTION

Genetic traces of selection

The history of human evolution is written in the genome, and researchers have now pinpointed gene variants that may have helped humans to adapt to their environment.

Sharon Grossman and Pardis Sabeti of the Broad Institute in Cambridge, Massachusetts, and their colleagues used a computational tool to root out single-letter changes in the human genome that have been under selective pressure in the past 50,000 years. They found that one such variant seems to temper the immune response to certain types of bacteria, and may have been selected for when humans were exposed to high levels of infections.

In a separate paper, Sabeti and her co-workers introduced another of the gene variants that they uncovered into mice. The resulting animals had thicker hairs and a higher number of active sweat glands — traits borne by humans with the same variant. However, whether extra sweat glands benefited early humans is a matter of speculation.

Cell 152, 703–713; 691–702 (2013)

REPRODUCTIVE BIOLOGY

Broken DNA in ageing eggs

The quality of egg cells declines as women age, probably partly because the cells' ability to repair DNA damage becomes impaired.

Kutluk Oktay at New York Medical College in Rye and his collaborators found that egg cells from older women have more DNA damage and lower expression of four DNA-repair

genes, than those from younger women.

When the authors silenced these genes in mouse egg cells using short interfering RNA molecules, the egg cells were more prone to DNA damage and death. When they overexpressed one of these genes, *BRCA1*, the old mouse eggs became as robust as the young ones.

These results indicate that the decline in egg-cell quality may be reversible and could lead to treatments that prolong female fertility, say the authors. *Sci. Transl. Med.* 5, 172ra21 (2013)



ATMOSPHERIC SCIENCE

Harder rains in a hotter climate

The amount of rain delivered in intense spurts, such as in thunderstorms, will probably increase more than that from drawn-out showers, as the climate warms.

As temperatures rise, so does the frequency of extreme precipitation events, but scientists have had trouble untangling which type of precipitation contributes most to these changes in weather. Jan Haerter of the University of Copenhagen and his colleagues used radar measurements and data from gauges, taken as often as every five minutes over many months, to

study how much rain fell over parts of Germany. Cloud observations were used to distinguish between the two types of rain.

The amount of rain delivered in extended showers increased with temperature at about the expected rate. However, the volume produced in intense bursts increased faster than the rise in the atmosphere's water-holding capacity, which also rises with temperature. Such erratic precipitation patterns may dominate in a warmer future.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo1731> (2013)

ZOOLOGY

Bright nights speed birds' lives

Exposure to low levels of artificial light at night can cause birds to become ready for reproduction earlier than those that experience dark nights.

Davide Dominoni and his colleagues at the Max Planck Institute for Ornithology in Radolfzell, Germany, attached miniature devices to wild European blackbirds (*Turdus merula*; pictured) to record the light levels that the birds experience in city

and forest environments. On the basis of these data, the authors exposed adult male blackbirds captured from both settings to either dark or low-light conditions at night. Birds kept in the brighter-night environment developed reproductive physiology nearly a month earlier and moulted sooner than their dark-dwelling counterparts.



SETH RESNICK/SUPERSTOCK/CORBIS

IMAGEBROKER/FLPA

The light level the authors used was 20 times lower than that produced by a streetlight, showing that even small changes can have an impact on animal development.

Proc. R. Soc. B 280, 20123017 (2013)

NEUROSCIENCE

Protein makes she sound like he

Levels of a brain protein could help to explain sex differences in some animal vocalizations, including in human language.

In a study of newborn rats, Michael Bowers at the University of Maryland School of Medicine in Baltimore and his colleagues identified a relationship between differences in the level of protein *Foxp2* in the male and female brain to the ultrasonic distress calls that pups make when they are separated from their mothers. Male pups — which have higher *Foxp2* levels in multiple brain regions — call more frequently and at lower pitches than their female siblings. When *Foxp2* levels were reduced in males and increased in females, these differences were reversed. The authors also found sex differences in the protein in human brain samples that correlates with earlier language acquisition by girls. *J. Neurosci.* 33, 3276–3283 (2013)

MOLECULAR BIOLOGY

Dance of DNA-binding proteins

Contortions of a DNA strand caused by the binding of a protein affect how a second protein binds — a phenomenon that may regulate gene expression.

Sunney Xie at Harvard University in Cambridge, Massachusetts, and his group used fluorescence microscopy to observe interactions between proteins and DNA at the single-molecule level. The authors

showed that one protein could stabilize or destabilize a second protein's attachment to DNA. The proteins do not interact directly. Rather, one widens or narrows grooves along the DNA helix, affecting the binding of another protein as far away as 30 base pairs. Such allosteric interactions also occurred in live bacterial cells.

Proteins that bind DNA are often involved in gene regulation, and these interactions may be a mechanism in its fine-tuning. *Science* 339, 816–819 (2013)

MEDICINE

Microbes make melamine toxic

Melamine, the chemical that achieved infamy in 2008 for the poisoning of children's milk in China, owes its toxicity in part to gut bacteria.

Wei Jia and Aihua Zhao of Shanghai Jiao Tong University in China and their group show that the toxicity of melamine in rats is reduced and its excretion increased if gut microbes are suppressed with antibiotics. They also show that the bacterium *Klebsiella terrigena*, which the authors had cultured from rat faeces, converts melamine to cyanuric acid *in vitro*. This acid — which forms crystals with melamine — is a key component of the kidney stones linked to both melamine-related kidney failure and death.

The authors suggest that melamine toxicity may depend on the composition of an organism's gut microbes. *Sci. Transl. Med.* 5, 172ra22 (2013)

EVOLUTION

Predictable bacterial diversity

Similar, and even identical, mutations underlie the metabolic diversification of independently evolving populations of the bacterium *Escherichia coli*.

Matthew Herron and Michael Doebeli at the

COMMUNITY CHOICE

The most viewed papers in science

CHEMISTRY

Hydrogen on demand

HIGHLY READ
on pubs.acs.org
in January

Silicon reacts extremely slowly with water to produce hydrogen gas, but 10-nanometre-wide silicon particles react 1,000 times faster.

Mark Swihart, Paras Prasad and their colleagues at the State University of New York at Buffalo report that 1 gram of powdered silicon can generate about 2 litres of hydrogen in about 45 seconds — a rate that is sufficient for systems to produce hydrogen fuel on demand from water.

The authors say that the process they used to generate the silicon powders — breaking up silane gas with lasers — can produce kilograms of powder in an hour, which may mean that the process has the potential to move beyond niche applications.

Nano Lett. 13, 451–456 (2013)

University of British Columbia in Vancouver, Canada, exposed *E. coli* to a mixture of glucose and acetate and found that competition for the carbon sources caused the bacteria, over about 1,200 generations, to evolve into two ecologically and metabolically different strains. This occurred in three independent evolution experiments. The authors sequenced the genomes of the three bacterial populations at 16 points during their evolution. They found many similar and a few identical mutations that underlay the evolution of diversity in the three experiments.

The findings suggest that this evolution is a predictable process that is driven by natural selection.

PLoS Biol. 11, e1001490 (2013)

ECOLOGY

Tough life in the tropics

The tropics host a much smaller number of invasive species than temperate regions, possibly because predation in the tropics is more intense.

Amy Freestone of Temple University in Philadelphia, Pennsylvania, and her colleagues conducted field



AMY FREESTONE

experiments off the coast of Connecticut and Panama. They allowed populations of non-native tunicates (pictured) — sessile marine creatures also known as sea squirts — to grow on plastic plates, blocked predators such as crabs and fish from reaching some of the plates, and looked at the effects of predation on species richness. Predated plates in the tropics had fewer non-native tunicate species than those in temperate areas.

The results generally support the idea that food webs in the tropics are more diverse and interconnected than those in temperate latitudes, making it harder for non-native species to establish themselves.

Ecology <http://dx.doi.org/10.1890/12-1382.1> (2013)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

US open-access bill

A bill that requires free public access to academic articles resulting from federally funded research was introduced into the US Congress on 14 February. If passed, the bill would order research agencies to give free access to papers no later than six months after their publication. Currently the US National Institutes of Health requires its research to be publicly accessible after 12 months. This is the fourth such bill on open access to be introduced into Congress. None has yet made it to the voting stage. See go.nature.com/wm4d77 for more.

Fusion support

The United States should continue to invest in research on inertial-confinement fusion, in which hydrogen isotopes are rapidly compressed until they fuse and produce energy, says a report released on 20 February by the US National Research Council. The document also

SOUND BITE

“If Congress won’t act soon to protect future generations, I will.”

US President Barack Obama warns in his State of the Union speech on 12 February that he will move to tackle climate change and reduce pollution if Congress fails to agree on a plan of action.



MARINA PUSTOVAYA/ITAR-TASS

Massive meteor strike

The Russian city of Chelyabinsk was rocked by the shock wave from a massive meteor that broke up high above the city on 15 February. More than 1,000 people were injured by flying glass as windows shattered throughout the city. Data from the Comprehensive Test Ban Treaty Organization show that the explosion was the

equivalent of around half a megatonne of TNT. That would correspond to a meteoroid around 17 metres across and weighing 10,000 tonnes. It is believed to be the largest object to strike Earth since a massive body broke up over the Tunguska river in Siberia in 1908. See go.nature.com/yxeiwq and page 281 for more.

calls for continued support for the National Ignition Facility at the Lawrence Livermore National Laboratory in Livermore, California, to help it to reach ignition — the point at which the isotopes produce more energy by fusing than is needed to bring them together.

New nuclear reactor

The Hongyanhe nuclear reactor in the northeastern Liaoning province of China fired up on 17 February — the first new reactor to start up in the country since it imposed a ban on new nuclear facilities after the 2011 disaster at Fukushima, Japan. The ban ended in October last year. China now has 16 reactors and a generating capacity of 12 gigawatts. It plans to be producing 58 gigawatts of nuclear power by 2020.

Gene-patent ruling

In a landmark court case, the Federal Court of Australia in Sydney ruled on 15 February that genes are patentable because the isolation of a human gene is the product of human intervention. The decision upholds Australian patent claims held by Myriad Genetics, a genetic diagnostics company based in a Salt Lake City, Utah, that is also facing patent challenges in the United States. The US case is likely to be heard before the Supreme Court later this year.

atomic structures of pathogens requiring level-3 biosafety containment, such as severe acute respiratory syndrome (SARS). The new lab, known as Crystal, is one of only two such synchrotron facilities in the world; the other is housed at the University of Chicago in Illinois. The lab opening was announced on 17 February and will aid the study of pathogens with a major global impact on human and animal health.

RESEARCH

CERN data boon

The computer banks at CERN, Europe's premier particle-physics laboratory near Geneva, Switzerland, have now recorded more than 100 petabytes (100 million gigabytes) of data. The data are held on 52,000

FACILITIES

Pathogen study

Britain's national synchrotron facility in Oxfordshire has opened Europe's first laboratory with facilities to study the molecular and

WILFRIED ROSENDAHL/REM
tape cartridges and on more than 17,000 disks at the lab; equivalent to roughly 700 years of high-definition-quality film. Most of the data — 75 petabytes — came from the lab's massive Large Hadron Collider (LHC), which discovered the Higgs boson last July. The announcement about the data came on 14 February, just as the LHC shut down for a two-year technical stop.

Childhood cancer

Research on childhood cancer should focus more on the needs of developing countries, say experts. In a series of papers published in *Lancet Oncology* on 20 February, the researchers say that deaths from childhood cancers have fallen in wealthy nations, with 94% of such deaths globally now occurring in low- and middle-income countries. They call on policy-makers to ease the way for clinical trials on innovative medicines for childhood cancers. See go.nature.com/wvxxom6 for more.

Medici skeleton

The first photographs of the skeleton of Anna Maria Luisa de' Medici, the last member of the powerful Florentine Medici family, who died in 1743, have been released. The skeleton of the princess, still adorned with her crown



(pictured), was exhumed in Florence, Italy, in October and is in better condition than researchers expected. DNA and isotope analysis of bone samples will try to uncover more about her lifestyle and cause of death. An exhibition on the Medici opened on 17 February at the Reiss Engelhorn Museum in Mannheim, Germany, which is collaborating with the University of Florence in the project. See go.nature.com/gws6g3 for more.

Go-ahead for iPS

The world's first clinical study that puts induced pluripotent stem (iPS) cells into humans was given a conditional go-ahead on 13 February. A review board at the National Institute of Biomedical Innovation in Osaka, Japan, gave the green light to a study to treat age-related macular degeneration, a condition that affects the retina and

can lead to blindness. Study leader Masayo Takahashi of the RIKEN Center for Developmental Biology in Kobe, Japan, now needs approval from the health ministry, which she is set to receive before the end of the next fiscal year in March 2014. See go.nature.com/bkvonk for more.

Hormones disrupted

Synthetic chemicals found in many household and industrial products can disrupt hormonal systems and could have significant health implications for people and wildlife, finds a report from the United Nations Environment Programme and the World Health Organization, published on 19 February. The study is the most comprehensive yet on endocrine-disrupting chemicals, which have been linked to reproductive abnormalities in young men and to breast cancer in women. It calls for more research on the chemicals' health impacts.

PEOPLE

Health survey

The head of a survey to measure the health effects of the 2011 Fukushima nuclear accident in Japan is stepping down. Shunichi Yamashita, a radiation health physicist from Nagasaki University, will leave his position as head of

COMING UP

22 FEBRUARY

The United Nations Environment Programme concludes its annual meeting at its headquarters in Nairobi by announcing a centre to help transfer climate-related technology to developing nations. www.unep.org/gc/gc27

24–26 FEBRUARY

IAP (formerly the InterAcademy Panel), a global network of science academies based in Trieste, Italy, holds a conference on poverty eradication and sustainable development in Rio de Janeiro, Brazil. go.nature.com/1cy96n

the health survey at the end of March 2013 when his contract expires. Yamashita had come under fire from local activists for claiming that the accident would cause few cancers. He will continue to work as a part-time consultant for Fukushima Medical University. A replacement has not been named. See go.nature.com/hekwld for more.

EVENTS

Chernobyl collapse

A massive section of the walls and roofing surrounding part of the ruined Chernobyl unit 4 nuclear reactor in Ukraine has collapsed. About 600 square metres of roofing and wall panels fell, probably as a result of a build-up of snow. No radiation leak was detected. The roofing was erected shortly after the accidental meltdown of the reactor in 1986. It will eventually be replaced by a large concrete arch, currently scheduled for completion around 2016.

➔ NATURE.COM

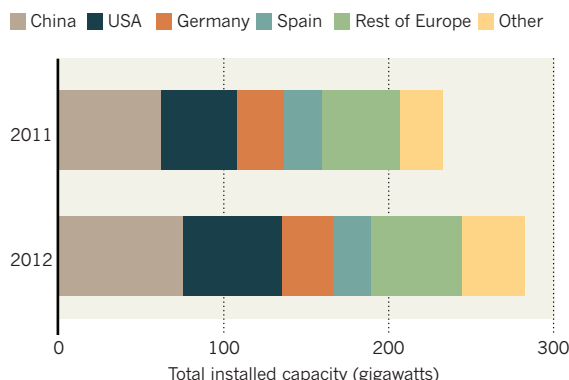
For daily news updates see: www.nature.com/news

TREND WATCH

The United States and China both installed around 13 gigawatts (GW) of new wind-energy capacity last year, according to provisional figures from the Global Wind Energy Council, the industry's trade association headquartered in Brussels. Of the world's total installed capacity of 282.4 GW, China has 75.6 GW, more than one-quarter (although not all turbines are connected to the grid). Worldwide, 5.4 GW (2% of total capacity) are offshore turbines, most of them in northern Europe.

MORE POWER FROM THE WIND

Global wind-power capacity grew by 18% in 2012, with most expansion in China and the United States.



NEWS IN FOCUS

GENOMICS A pause for breath in the genome-sequencing race **p.290**

COSMOLOGY Dark pursuits under a marble mountain in China **p.291**

CHINA Bountiful shale gas will be slow to dethrone coal **p.294**



MATERIALS Chemists churn up cleaner, greener cement **p.300**

DANIEL ACKER/BLOOMBERG/GETTY



Monsanto could face a loss of revenue from its herbicide-resistant soya beans.

BIOTECHNOLOGY

Seed-patent case in Supreme Court

Loss of patent control could rekindle 'terminator' technology.

BY HEIDI LEDFORD

A technology called a 'terminator' was never going to curry much favour with the public. But even Monsanto, the agricultural biotechnology giant in St Louis, Missouri, was surprised by the furore that followed when it patented a method for engineering transgenic crops to produce sterile seed, forcing farmers to buy new seed for each planting. In 1999, Monsanto's chief executive pledged not to commercialize terminator seeds.

The concept, if not the technology, is now gaining traction again. This week, the US Supreme Court hears arguments that pit

Monsanto against 75-year-old Indiana soybean farmer Vernon Hugh Bowman, who used the progeny of Monsanto seeds to sow his land for eight seasons. The company says that by not buying seeds for each generation, Bowman violated its patents. If Bowman wins — and observers say that is not out of the question — the decision could make it harder for biotech firms to enforce patents on engineered organisms, from seeds to microbes, prompting them to revisit terminator-like technology.

"If I were at Monsanto and I learned that patents are not available to protect my soybeans, I would think of some kind of technological fix," says Christopher Holman, an

intellectual-property specialist at the University of Missouri-Kansas City School of Law. Indeed, some synthetic-biology companies, concerned about policing illegal, copycat proliferation of their technology, are already working on terminator-like safeguards.

Bowman was a regular customer for Monsanto's herbicide-resistant soya beans for his main crop, but bypassed the company by purchasing seed for a late-season crop from a grain elevator known to contain Monsanto's transgenic seed. In 2007, Monsanto sued him. As the case climbed through the court system, it grew from a simple contract violation to a challenge of the idea that companies can use patents to limit the offspring of naturally 'self-replicating' technologies. The lower courts sided with Monsanto, and many were surprised when the Supreme Court took up the appeal.

Even if the Supreme Court does not overturn the lower courts' rulings outright, biotechnology firms are bracing themselves for clarifications of patents on self-replicating inventions. Without that protection, companies say, they have little recourse to prevent someone from buying seed — or a cell culture or a transgenic animal — and using it to generate thousands more to sell again at a fraction of the original price. "Once you have sold the first seed, you are done," says Hans Sauer, deputy general counsel for intellectual property at the Biotechnology Industry Organization, a lobby group in Washington DC.

Early patents on 'gene-use restriction technologies' — later rebranded as 'terminator' technology by activists opposed to them — described a genetic modification that switched on production of a toxin that would kill off developing plant embryos. The result: a seed that could be harvested for food but would not produce offspring. The controversial proposal raised concerns that it would make farmers dependent on industry for their livelihood.

There are alternatives to making sterile seeds (see 'Terminator, the sequel'). One tactic would be to switch off the transgene of interest in seeds, so that they could grow into new plants but would not pass on the benefits of the engineered trait. Another approach is to place the transgene under the control of a switch that must be activated by a proprietary chemical. That would give companies control over the engineered trait by forcing buyers to return each year to purchase the chemical.

That is the strategy of Ginkgo BioWorks, ►

► a four-year-old synthetic-biology company in Boston, Massachusetts, that develops ‘made-to-order microbes’ to churn out marketable chemicals. Founder Jason Kelly says that the company plans to charge customers on the basis of how much they use the microbes. For accurate billing and theft protection, Ginkgo needs to control that use, so it is developing what Kelly calls a ‘gene-guard’ technology: a genetic tweak that makes production of the desired chemical dependent on a proprietary additive, supplied by Ginkgo, in its fermentation medium. The approach could even be used in nanotechnology, by making engineered nanobots that are dependent on a proprietary raw material.

In a strange twist of fate, terminator technology has begun to look more appealing to environmentalists. Organic farmers want ways to keep genetically engineered crops from contaminating their fields, and food-safety groups are concerned about contamination of food crops with products from a new generation of crops engineered to produce chemicals or pharmaceuticals. By ensuring that genetically modified plants survive for only one planting, “that technology would have alleviated a lot of environmental concerns”, says Holman.

Approaches dependent on switching will take considerable research, however; companies reported technical challenges with the first generation of terminator techniques. Patents owned by Monsanto required the insertion of three different genes into the plant genome. Monsanto says it is currently not researching the techniques, and other companies are hoping that they will not have to. “Perhaps these technologies could provide new ways to protect investments,” says Brett Lund, former head of intellectual property for the biofuels group of Syngenta, an agri-giant headquartered in Basel, Switzerland. “But the easiest and best way is through our patent system.” ■

TERMINATOR, THE SEQUEL

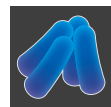
As the US Supreme Court hears a controversial gene-patent case, biotechnology companies are once again considering ways to limit naturally self-replicating technologies.



SEEDS

‘Terminator’ technology produces sterile seeds.

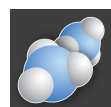
Status: Politically toxic.



SYNTHETIC BIOLOGY

‘Gene guard’ needs proprietary supplement in growth medium.

Status: Under development.



NANOTECHNOLOGY

Requirement for proprietary supplement to ‘grow’.

Status: Hypothetical.



TRACEY GRIFFITHS

Single-cell genomics is allowing fertility clinics to screen embryos for abnormalities more cheaply.

GENOMICS

Gene sequencing leaves the laboratory

Maturing technology speeds medical diagnoses.

BY ERIKA CHECK HAYDEN

The steep fall in the cost of sequencing a genome has, for the moment, slowed. Yet researchers attending this year’s Advances in Genome Biology and Technology (AGBT) meeting in Marco Island, Florida, on 20–23 February are not complaining. At a cost as low as US\$5,000–10,000 per human genome, sequencing has become cheap and reliable enough that researchers are not waiting for the next sequencing machine to perfect new applications in medicine.

Two recent beneficiaries: embryo screening after *in vitro* fertilization (IVF) and metagenomic medicine, which entails sequencing many different microbes en masse and then teasing out individual genomes to diagnose which ones are helping or harming human health. “The field is maturing,” says geneticist Jay Shendure at the University of Washington in Seattle.

At last year’s AGBT meeting, Oxford Nanopore, based in Oxford, UK, electrified researchers by unveiling details of a machine that could sequence DNA at unprecedentedly low cost and high speed — but that machine has not yet arrived on the market. This year, no splashy new technologies are expected. Instead, the HiSeq machine, made by Illumina, based in San Diego, California, has tightened its grip on the top end of the market (see ‘Reasonably priced genomes’).

Borrowing a metaphor from computer-circuit development, quantitative biologist Michael Schatz at Cold Spring Harbor Laboratory in New York says that after many “tick” years of dramatic technological innovation, the field is now in a “tock” period as researchers fine-tune the technology and expand it into applications.

One application depends on single-cell genomics, a focus of this year’s meeting. To sequence a genome from a single cell, the DNA has to be amplified, a process that can introduce errors into the sequence. But new amplification techniques now allow accurate sequences to be obtained from individual cells, including the precious cells extracted from human embryos.

At the meeting, Dagan Wells at the University of Oxford, will describe how the technique can be used to select which of the embryos created by IVF has the best chance of developing into a healthy baby. Fertility experts already screen embryos for genetic abnormalities by extracting a single cell from each eight-cell embryo and running a variety of tests to detect chromosomal abnormalities, but this can cost around a thousand dollars per IVF cycle.

Instead, Wells extracts DNA from the sampled cell, amplifies it and sequences a part of the genome using the Personal Genome Machine, a fast sequencer made by Ion Torrent in Guilford, Connecticut, a subsidiary

of Life Technologies in Carlsbad, California. Within a day, Wells can detect duplications or deletions of entire chromosomes — the abnormalities seen most often early in development. Several women are now pregnant with embryos selected using this technique, and a larger clinical trial is planned for later this year. Wells says that the low cost — as much as 50% less than current methods — should open up testing to more women.

Other researchers are refining existing technologies to increase the average length of the DNA strand that can be 'read' from the sequencer. The devices that dominate the field sequence short fragments of DNA, no more than a few hundred base pairs, and overlook some portions of the genome. Longer reads are needed, for example, to distinguish between closely related microorganisms — an important requirement for researchers studying whether overabundance of some microbial species in the gut or vagina make patients more susceptible to problems ranging from diarrhoea to premature birth.

Late last year, Illumina bought Moleculo, a company based in San Francisco, California, whose technology extends reads by changing the way that DNA is prepared for sequencing and by applying new analytical software to the readout. These tricks can coax reads thousands

REASONABLY PRICED GENOMES

Although no reports of big innovations in DNA sequencing are expected at a major conference this week, the current cost and capabilities of the technology now make medical applications worthwhile.

| Name | Machine cost | Read length (bases) | Cost per megabase |
|--------------------|--------------|---------------------|-------------------|
| Illumina MiSeq | US\$125,000 | 500 | 14–70 cents |
| Illumina HiSeq | US\$690,000 | 300 | 4–5 cents |
| PacBio RS | US\$695,000 | 4,575 | \$2–17 |
| Ion Torrent PGM | US\$49,000 | 400 | 60 cents–\$5 |
| Ion Torrent Proton | US\$224,000 | 200 | 1–9 cents |

of bases long out of Illumina machines — an order of magnitude greater than previously possible. The machines produced by Pacific Biosciences in Menlo Park, California, also achieve read lengths that routinely reach into the thousands of bases, thanks to an engineered polymerase. Customers say that the company seems to have overcome early problems with reliability.

In market share, Pacific Biosciences still has a long way to go to catch Illumina or Life Technologies, which this month said that it expects sales of the Ion Torrent sequencing system to boost revenue by up to 5% in 2013. Corporate takeovers are a topic of fevered speculation in the sequencing world. Rumours that Life Technologies' sequencing business may be

acquired sent its shares soaring to their highest price ever. And Illumina recently fended off a hostile takeover bid by Swiss company Roche in Basel, which formally dropped its offer a month ago. But the fundamental landscape has not changed much.

Any reshuffling of the major players will have to wait for the field's next 'tick' cycle, whether it takes the form of Life Technologies' Ion Proton machine — which is just hitting labs and which the company says will eventually deliver a \$1,000 genome in a day — or Oxford Nanopore's GridION and MinION machines, which scientists still haven't seen. In the meantime, says Shendure, "this year's meeting will be less glitz and more science, and that's great". ■

SOURCE: THE COMPANIES; TRAVIS GLENN

PHYSICS

Dark-matter hunt gets deep

China launches world's deepest particle-physics experiment — but it joins a crowded field.

BY EUGENIE SAMUEL REICH

More than 1,000 metres underground, physicists have set traps of liquid xenon to catch their prey: hypothetical particles of dark matter that might very rarely interact with ordinary matter as they drift through Earth. With construction costs on the order of US\$10 million each, such experiments are a relatively cheap way to work out the composition of 85% of the matter in the Universe. But does the world really need four of them?

Ongoing experiments in Italy, the United States and Japan are now being joined by a fourth in China, called PandaX (see 'Dark and deep'). Installed in the deepest laboratory in the world, 2,500 metres under the marble mountain of JinPing in Sichuan province, PandaX will this year begin monitoring a tank containing 25 kilograms of xenon. The team hopes to scale the tank up to 1 tonne by 2016, which would mean that the experiment had developed more quickly than any other dark-matter search. "We



A conveyor belt removes rock from JinPing laboratory, a 2,500-metre-deep dark-matter experiment site.

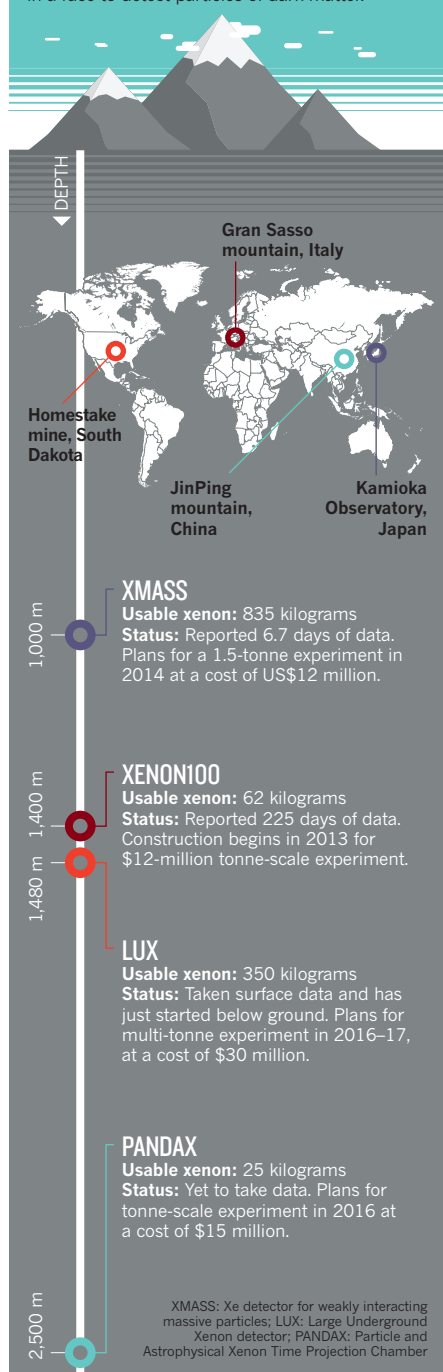
want to demonstrate that world-class research in dark matter is possible in China," says Xiangdong Ji, a physicist at Shanghai Jiao Tong University in China and a spokesman for PandaX.

Dark-matter researchers in the West are excited by the ambition of the project, but some question the duplication of effort. "Spending all our money on different direct-detection ►

WOLFGANG LORENZON

DARK AND DEEP

Shielded from cosmic rays by the bedrock, four experiments are using giant tanks of liquid xenon in a race to detect particles of dark matter.



► experiments is not worth it,” says Stefan Funk, an astrophysicist at the SLAC National Accelerator Laboratory in Menlo Park, California, who admits that he is biased — he looks for dark matter indirectly, by watching the heavens for a possible γ -ray signal.

The proliferation of xenon experiments reflects infighting among the dark-matter hunters, as well as national ambitions to be first to answer one of physics’ most compelling questions. Observations of the rotation of galaxies and of the cosmic background radiation — the afterglow of the Big Bang — suggest that some 85% of matter in the Universe exerts gravitational pull, but does not radiate light. Leading theories suggest that such dark matter comes in the form of weakly interacting massive particles (WIMPs). If they exist, these would occasionally collide with nuclei of ordinary matter, so they could in principle be directly detected in a large tank of an atomically dense material such as xenon, which produces light when a nucleus recoils.

Such experiments are placed underground to minimize interference by cosmic rays that can mimic WIMPs. So far, they have seen nothing. The most progress has been made by XENON100, an experiment near L’Aquila, Italy, which has not detected a WIMP signal in 225 days of running time so far, ruling out the existence of heavier and more strongly interacting particles. In a few months, the XENON collaboration will start to scale up its detector to more than 1 tonne of xenon, making it 100 times more sensitive.

Enter PandaX, which was born when China’s National Natural Science Foundation and Ministry of Science and Technology funded a team that included members of the XENON collaboration. They set up an experiment of their own, deep within JinPing mountain, in a small laboratory that opened in 2011 after just two years of construction. PandaX hopes to rival XENON100’s current sensitivity for lighter WIMPs by the end of this year, although Ji acknowledges the challenges of working in a remote area of central China while relying on overseas suppliers for many parts. Just last week, the team was unable to cool down its experiment because of a delay in the delivery

of liquid nitrogen. But because PandaX is more than a kilometre deeper than the other experiments, it needs less shielding from cosmic rays, which will make it easier to scale up.

China, together with a consortium of universities including some in the United States, is spending about \$15 million to build PandaX, but the team would like more international partners. In January, PandaX applied to the US Department of Energy (DOE), which has not set up an agreement to work with the Chinese foundation, but the US agency declined. Instead, the DOE is spending its money on another group that split off from the XENON collaboration: the Large Underground Xenon (LUX) experiment at the Homestake gold mine in South Dakota.

Homestake was once meant to be the site of a major underground laboratory. But the lab’s main funder, the US National Science Foundation (NSF), dropped the idea in 2010 (see *Nature* <http://doi.org/bcbr7z>; 2010), leaving the DOE to run it alone and bear the whole cost of the constant pumping required to prevent the mine from flooding.

LUX might be the most vulnerable of all the xenon experiments, says Juan Collar, a physicist at the University of Chicago in Illinois who looks for lighter WIMPs with a number of smaller experiments. “The funding situation is really terrible,” he says. The DOE has indicated that it plans to reduce the number of large dark-matter experiments it funds in October, so LUX could face the axe. US contributions to XENON100 are funded through the NSF, so it won’t be subject to the winnowing process.

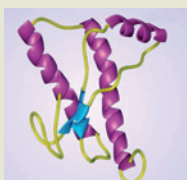
Rick Gaitskill, a physicist at Brown University in Providence, Rhode Island, and spokesman for LUX, says that the United States needs its own deep underground lab to maintain its dark-matter expertise. “Why would we give up that leadership position and move all that experience overseas?” he asks.

But Ji, who now has access to perhaps the best underground lab in the world, says that some consolidation is inevitable. The international community is unlikely to support more than two xenon experiments with multi-tonne detectors, he says, and the United States will need to choose which effort to back. In the meantime, he says, it is not a bad idea to have many groups working to improve the technology. “That will help build the ultimate dark-matter experiment.” ■

“Spending all our money on different detection efforts is not worth it.”

MORE ONLINE

TOP STORY



Well-folded prion proteins play ‘good’ role in brain development
go.nature.com/bc47t5

MORE NEWS

- The last of the Medici line may not have died of syphilis after all go.nature.com/gws6g3
- Tracking complex networks from a few nodes go.nature.com/o92aja
- Supernova shockwaves produce cosmic rays go.nature.com/phyf5

ENVIRONMENT



Anti-anxiety drugs increase aggression in river fish
go.nature.com/d9nloa

BENT CHRISTENSEN



Italian prime minister Mario Monti (left) with virologist Ilaria Capua, newly minted political candidate.

ITALY

Reforms at stake in Italian election

Italy's researchers want change they can believe in.

BY ALISON ABBOTT

Virologist Ilaria Capua spent the Christmas holiday planning her exit strategy from Italy, where she is currently director of research at the IZSVE (the Experimental Institute for the Prevention of Animal Diseases), a public veterinary institute in Legnaro, in north-east Italy. Internationally renowned for her work on the H5N1 avian flu virus — and more patient than most with Italy's notoriously obstructive bureaucracy — she finally snapped last November after administrators at her institute blocked her from accepting an offer of generous lab space in a new research centre in a hospital in Padua.

Then, out of the blue on 6 January, she got a call from Mario Monti, Italy's outgoing prime minister, who was appointed on 12 November 2011 to head a caretaker government that has attempted to tackle Italy's economic problems and debt. He offered her a safe seat in the parliament if she agreed to stand for his newly created party Civic Choice in the national elections, which take place this weekend. Capua, until now dedicated single-mindedly to her research, agreed. "I hope I can help promote the meritocracy that we desperately need in Italy," she says.

All the major parties, including ex-prime minister Silvio Berlusconi's People of Freedom party and the populist Five Star Movement party, have acknowledged the importance of science. But Monti has gone further: nearly a third of the candidates he has chosen are scientists or academics rather than career politicians — a reflection of his belief that only technical expertise can save Italy from decline. Although the next government is likely to be led by Pier

Luigi Bersani's centre-left Democratic Party, a coalition called 'With Monti for Italy' (*Con Monti per l'Italia*), including Civic Choice and two other centre and centre-right parties, may become a junior partner.

The country's researchers are counting on the next government to reverse the trend of annual budget cuts for universities and research institutes. They also hope it will allow a series of radical reforms to Italian universities and research institutes, introduced over the past three years, to go forward without interruption.

The sweeping reforms were designed to clean up bad habits such as nepotism and to cut the bureaucracy that hampers everything from academic appointments to buying lab equipment. The research institutes of all 12 national research agencies have been reorganized and the practice of government appointment of their presidents abandoned, so that candidates are appointed on the basis of merit rather than by political fiat.

New rules for academic appointments require universities to select their professors from a national list of qualified candidates (non-Italian candidates will not need to qualify for the list). PhD courses will have to be approved by a national committee and the organizers must prove that they have sufficient teaching staff. All public research money will now be distributed only after a peer-review panel, which must include some foreign researchers, has vetted the applications.

Central to the reforms has been the creation of a national evaluation agency, ANVUR, which launched a massive exercise to evaluate universities and research agencies in May 2011. The outcome should determine how at least 13% of government funding to universities is distributed, and some researchers, especially in the humanities, have criticized the metrics used.

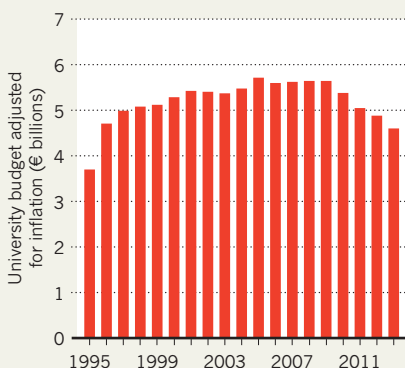
But the noisy resistance to ANVUR has raised concerns that politicians will be tempted to tinker with the reforms or stall them. "The worse thing for us would be that the reforms would be stopped or new ideas forced upon us," warns chemist Dario Braga, dean of research at the University of Bologna. "We are in transition and need to be left in peace."

Crippling budget cuts have amplified the tensions. University budgets have fallen by more than 10% since 2009, to €6.7 billion (US\$9 billion), and research-agency budgets have stagnated. To save money, the government has allowed only around one in five retiring professors to be replaced since 2008, reducing the number of academic staff from more than 62,000 to less than 55,000 in 2012. Competitive research money is minimal: just €38 million for basic research across all disciplines this year. "This revolution needs to be funded with more money, not less," says Braga. ■

Additional reporting by Nicolò Musner.

IN DECLINE

University funding in Italy has been cut in real terms since 2005.





Technicians inspect a conventional gas well in China's Tarim basin, which is also rich in shale gas.

ENERGY

China slow to tap shale-gas bonanza

Geology and infrastructure could impede development.

BY JEFF TOLLEFSON

After more than a decade of spectacular growth fuelled by coal, China finds itself sitting on a bonanza of shale gas. Its reserves are the world's largest, beating even those of the United States. But developing this vast resource won't be easy, as a bidding last month for shale-gas leases made clear.

"The resource is huge," says Jane Nakano, a fellow of the Energy and National Security Program at the Center for Strategic and International Studies in Washington DC. "But the shale deposits are more complex than ours, and the above-ground challenges are probably even larger" than the geological ones.

To offset some of the coal use that contributes to its status as the world's largest greenhouse-gas emitter, China wants to boost natural gas from around 4% of the country's energy mix to 10% by 2020. Much of that gas will be imported. But in March 2012, the Chinese government estimated the country's reserves at 25 trillion cubic metres, and an earlier estimate from the US Energy Information Administration was even larger. China's leaders resolved to boost annual shale-gas production from near zero today to at least 60 billion cubic metres by 2020. The United States, by

comparison, produced more than 150 billion cubic metres in 2010.

There, the abundant, cheap gas has displaced coal as a fuel for power plants, contributing to a nearly 4% fall in the country's fossil-fuel emissions in 2012. If China could repeat that success, the emissions reductions could be globally significant. But its shale-gas auction — only the second so far — has bolstered scepticism. China's Ministry of Land and Resources awarded leases in 19 areas, mostly in the nation's central Sichuan Basin. Analysts were surprised to see national oil and gas companies,

such as PetroChina and Sinopec, lose out to state-owned coal and utility companies, as well as to local government entities that have no expertise in the oil and gas arena.

Nakano says that the national oil firms may be playing it safe and did not truly compete to win. Price controls on natural gas may have reduced their appetite for risk, she says, and they have little experience with the hydraulic fracturing needed to release gas from shale.

Shu Jiang, a petroleum engineer who worked in China before moving to the University of Utah in Salt Lake City, is more optimistic, pointing out that major oil and gas companies are investing in shale-gas wells in areas already leased for conventional oil and gas development. He says that early results from the Sichuan basin are promising and that "China's vast shale resources will be extracted".

That is unlikely to happen quickly, however, says Julio Friedmann, chief energy technologist at the US energy department's Lawrence Livermore National Laboratory in California. "In the United States, it took 60 years and 200,000 wells" to lay the groundwork for the shale-gas revolution. China has drilled fewer than 100 wells, and its geology is different. Many of the Chinese shale formations have a high clay content, for instance, which makes them more pliable and less apt to fracture. Many are also deeper. "We simply have no idea about whether or not the geology is going to produce," Friedmann says.

The United States also has a unique abundance of small independent oil and gas companies with a tradition of risk-taking. China does not have such a 'wildcatter' culture, nor does it have the same mechanisms for developing and sharing geological data. Even a simple shortage of drilling rigs could slow things down.

Many expect that shale gas would at first be used in China as a feedstock for the chemical and fertilizer industries, reducing the use of gasified coal. In theory, that could lower the nation's annual carbon dioxide emissions by as much 100 million to 150 million tonnes, or roughly 1–1.5% of the nation's cumulative carbon emissions, Friedmann estimates.

Some of the climate benefits would be lost, however, if the wells leak methane — a potent greenhouse gas — and a shale-gas rush would raise the usual concerns about air and water quality. Water is scarce in the gas-rich Tarim basin (see 'Promising grounds'), whereas the Sichuan basin is heavily populated.

Groups such as the Clean Air Task Force, in Boston, Massachusetts, are encouraging companies and officials to adopt drilling practices that will minimize environmental impacts. Jonathan Banks, the force's climate-policy coordinator, says that the goal is to help China avoid mistakes made in the United States.

"China is going to be able to leapfrog over some of the stages that the United States went through," Banks says. "We are pushing to make sure that it is leapfrogging the environmental impacts as well." ■ [SEE COMMENT P.307](#)

PROMISING GROUNDS

China's shale-gas basins could hold 25 trillion cubic metres of the gas, but few wells have been drilled.



► **NATURE.COM**
Explore the world's
energy resources:
go.nature.com/3em370

SOURCE: EIA

FUNDING

Oil money takes US academy into uncharted waters

Venerable government adviser will fund grants with half-billion-dollar windfall.

BY HELEN SHEN

Breaking its 150-year-old tradition as a prominent scientific advisory group for the US government, the National Academy of Sciences (NAS) is quietly preparing to launch a major research-grant programme. The programme will be part of a US\$500-million, 30-year initiative to study environmental science and human health in the wake of the 2010 oil spill in the Gulf of Mexico.

The academy, based in Washington DC, was awarded the funds under two legal settlements stemming from the Deepwater Horizon oil spill, which spewed more than 600 million litres of oil into the gulf. With the awards, the NAS joins the small club of non-governmental organizations poised to support much of the research in the gulf in coming years (see 'Oil money everywhere'). But the timetable for the spending, and the research areas to be supported, are uncertain. "We're trying to get our plans in place," says NAS president Ralph Cicerone. "We're not ready yet."

Over the next five years, British oil-and-gas giant BP will give the NAS \$350 million, and Swiss oil-rig contractor Transocean will add \$150 million — money that officials at the NAS say they were surprised to receive. The US Department of Justice, which organized the settlements, said in a statement that the NAS was chosen for its institutional independence. The first \$5 million is scheduled to arrive by 29 April and larger instalments are slated for coming years. Cicerone says that "a fair amount" of the money will support external research projects — an unusual role for the institution, which mostly prepares scientific reports for government agencies.

Some scientists worry that the academy lacks experience in competitive peer review and grant oversight. "I don't know what entity within the national academy would be involved," says James Carr, a fish biologist at Texas Tech University in Lubbock. "We haven't heard anything about what the mechanism is going to be."

Cicerone acknowledges that setting up the research programme will lead the NAS into new territory, but points to the academy's experience in convening experts to conduct and peer review reports. "There's really nothing

OIL MONEY EVERYWHERE

Penalties and reparations from the 2010 Deepwater Horizon spill have enriched three non-governmental organizations that support research.

| Organization (founding date) | Source | Amount | Objectives | Funding start/end |
|------------------------------|-----------------------------------|------------------|---|---------------------|
| GoMRI (2010) | BP donation | US\$500 million | Oil dispersion and degradation studies; ecological and health effects | 2010–20 |
| NFWF (1984) | BP and Transocean plea agreements | US\$2.55 billion | Restoration and risk-reduction projects; construction of barrier islands | 2013–until depleted |
| NAS (1863) | BP and Transocean plea agreements | US\$500 million | Environmental protection and human health studies; education and environmental monitoring | 2013–43 |

GoMRI, Gulf of Mexico Research Initiative; NFWF, National Fish and Wildlife Foundation; NAS, National Academy of Sciences

so different there," he says. The academy is also trying to learn from organizations that already have a head start on distributing grants in the area, Cicerone says.

The National Fish and Wildlife Foundation in Washington DC, which received \$2.55 billion from BP and Transocean under the plea agreements, is focusing on recovery projects — including barrier-island construction off the Louisiana coast and river-diversion projects along the Mississippi River. By contrast, the Gulf of Mexico Research Initiative (GoMRI) — a \$500-million, 10-year competitive grants programme established by BP in the months following the oil spill — supports basic science. Modelled after the National Science Foundation (NSF), GoMRI has funded studies on the distribution and degradation of oil and chemical dispersants, as well as their ecological impacts. Rita Colwell, the chair of the GoMRI research board and a former director of the NSF, says that her organization has been holding discussions with the NAS. "We're willing to share what we've learned."

The NAS has yet to decide how it will balance its portfolio between basic and applied research, ecological and human-health studies, and short-term and long-term projects. Maureen Lichtveld, an environmental-health researcher at Tulane University in New Orleans, Louisiana, would like to see more funding for human health, which she says has received less support from GoMRI than ecological research. Lichtveld hopes that NAS

"Once the cheques start showing up, there will be an enormous amount of pressure to spend."

funds will enable her to expand her studies of women and children who have eaten seafood from the Gulf. Samantha Joye, a biogeochemist at the University of Georgia in Athens who studies sea-floor microbes and invertebrates, views the academy's 30-year mission as an opportunity to support extended — and continuous — ecological monitoring projects.

Cicerone says that the grants programme will complement the other funding sources. In January, NAS representatives spoke to Joye and several other scientists at a GoMRI-sponsored conference in New Orleans to identify research areas that have lacked support.

Steven Murawski, a fisheries biologist at the University of South Florida in St Petersburg, says that there is no shortage of things to do. He says that the academy faces a daunting task — not just in identifying priorities but also in disbursing funds quickly, while the effects of the oil spill linger. "Once the cheques start showing up, there will be an enormous amount of pressure to spend." ■

CORRECTIONS

The map in the story 'Quake fears rise at Japan's reactors' (*Nature* **494**, 14–15; 2013) wrongly suggested that only a single reactor, rather than the whole plant, is under inspection at Oi.

The News story 'Company offers portable peer review' (*Nature* **494**, 161; 2013) wrongly stated that the European Molecular Biology Organization's open peer review was not anonymous — the reviews are made public but are anonymized.

➔ **NATURE.COM**
For more on the Deepwater Horizon spill, see:
nature.com/oilspill

As the worm turns

With the help of a tiny worm, Cornelia Bargmann is unpicking the neural circuits that drive eating, socializing and sex.

BY STEPHEN S. HALL

Male sexual dysfunction is never pretty, even in nematodes. In normal roundworm courtship, a slender male will sidle up to a plump hermaphrodite, make contact, and then initiate a set of steps leading up to insemination: a sinuous backwards motion as he searches for the sexual cleft, a pause to probe, and finally the transfer of sperm. The whole business is usually over in a couple of minutes. "It's very slithery, and affectionate," says Cornelia Bargmann, who has been observing the behaviour of this particular worm, *Caenorhabditis elegans*, for 25 years.

Last October, scientists in Bargmann's laboratory at the Rockefeller University, New York, reported the discovery of a gene that seems to be crucial to successful mating. Disrupting the action of this gene causes male sexual confusion of almost epic pathos: nematodes with certain mutations poke tentatively at an inert hermaphrodite, making confused, fruitless curlicues around the potential mate. Occasionally the mutant male succeeds, but often he literally falls off the job and begins the search anew for a mate. Jennifer Garrison, a postdoc of Bargmann's who tracked the behaviour of these males, just shakes her head as she replays the scene on her computer screen. "Really sad," she says.

There are two punchlines to this story of thwarted invertebrate mating. One is the charming squeamishness with which Bargmann describes it, hesitating at words such as "vulva" and "spicule" and other anatomical gewgaws of roundworm reproduction. "As a well-brought-up Southern girl," she says with a laugh, "it's still difficult to talk about this!"

The other is scientific, supporting Bargmann's long-standing conviction that studying these deaf, part-blind, transparent creatures, which resemble nothing so much as wriggling specks of lint, could nonetheless yield enormous insight into how a nervous system creates behaviour. Since the 1980s, Bargmann and her colleagues have systematically explained the means by which worms taste and smell, exhibit social behaviours such as feeding in groups and explore their surroundings. She and her colleagues have parsed these behaviours down to the genes and circuitry of the neural connections that drive them. Just as studies of the fruitfly laid bare the basic principles of development and studies of yeast revealed the rules of the cell cycle, Bargmann believes that the simple nematode is revealing basic secrets about how animal nervous systems — including those of humans — translate sensory information into fundamental behaviours. "What are the most basic behaviours that

every animal has to show and every animal has to solve?" she asks. "You can basically say that the three would be hunger, fear and reproduction. None of those things got invented last Saturday night!"

The evolutionary strength of that argument grew last year, when Bargmann and her colleagues published the experiments with the mutant males¹. The mutation, they reported, is in a roundworm gene that they dubbed *nematocin*, which codes for a small peptide that influences multiple neurons and is a biochemical cousin to oxytocin and vasopressin, two hormones that play key parts in mammalian reproductive behaviour. Put another, Darwinian, way, the sexual confusion in mutant nematode males is tied to a molecule that seems to have been conserved in the nervous systems of animals at least since worms separated from vertebrates an estimated 600 million years ago. "Oxytocin and vasopressin are kind of at the top of the hierarchy of human neuropeptides in the brain," says biologist Scott Emmons of Albert Einstein College of Medicine in New York. "And you look in the worm and, lo and behold, you see the same thing, which is quite striking."

Although there had been some initial experimental forays into the behaviour of worms, says Paul Sternberg, who studies the nematodes at the California Institute of Technology in Pasadena, Bargmann took a more systematic, rational approach to dissecting its circuitry. "She made a major impact by taking that risk, and going broad and deep. She committed her life to this, and it's worked out beautifully."

SPEAKING THE LANGUAGE

Bargmann, now 51, grew up in Athens, Georgia, but steeped in European culture. Both her parents were born in Germany. Her father spoke to her in English; her mother in German. "Apparently, until I was 4 or 5 years old, I didn't realize that each of us spoke a different language when we were talking to each other," she says. Her fluency threw open a world of German-language books, including semi-popular accounts of animal behaviour written by Konrad Lorenz² and Karl von Frisch³. "My mother had these books," Bargmann recalls. Decades later, some

of these pioneering ethology works are almost compulsory reading in the Bargmann lab.

Bargmann says that her true interest in science, however, dates back to an adolescent prank in school. "We were taking this Earth-science class,

► NATURE.COM

To watch worm mating in action, see: go.nature.com/rkvrty



and the teacher told us that sodium was a metal and that if you placed it in water, it would burn," she says. "We were just electrified." With several classmates, Bargmann conspired to steal all the sodium from the school lab and test the hypothesis. "The sodium was flushed down the [boys'] toilet, which was blown off the wall!"

"This was literally the first time that science really struck me as something fun and exciting," Bargmann recalls. But she wasn't there to witness the explosion. "The ethics of the eighth grade is that you could be involved in stealing the sodium and you could be involved in planning the event," she says. "But there was no way that a girl was going to go into the boys' locker room." After a pause, she adds, with a smile, "It was a moment of weakness that I regret to this day."

Bargmann has put that regret to good use, crashing through one door after another ever since. She went on to become a biochemistry major (and class valedictorian) at the University of Georgia in Athens, before heading for graduate studies to the laboratory of Robert Weinberg, a molecular biologist at the Massachusetts Institute of Technology (MIT) in Cambridge. "Still waters run deep," says Weinberg, recalling the arrival of the soft-spoken Bargmann in 1981. "In the beginning, she didn't say much and was generally quiet, but I soon realized that she had a superior brain."

At the time, the Weinberg lab was using the newly minted tools of molecular biology to tease apart the mechanics of oncogenes — genes that, when mutated or hyperactive, trigger the unbridled cellular proliferation that drives cancer. Bargmann isolated and sequenced an oncogene called *neu* from a rat tumour⁴. Later, researchers would discover that some malignancies, notably breast cancer, express the same gene, known in humans as *HER2*. In the 1990s, the Californian biotechnology company Genentech in South San Francisco developed a drug to target breast cancers that overexpress this gene, and the resulting monoclonal antibody, trastuzumab (Herceptin) has since been used to treat nearly one million people with cancer. Bargmann says that she derives "immense personal satisfaction", but no significant royalties, from this early work.

Despite the heady start in a hot and highly visible field, Bargmann felt intimidated by the intellectual firepower of the people already in it. "It just wasn't so clear to me what I would do that was so different from what people like Weinberg and Harold Varmus were already doing," she says. So after receiving her PhD in 1987, she made a strategic — and quietly ambitious — decision to switch her area of research. Bargmann had always been fascinated by the neuroscience of behaviour, and the moment seemed right to bring molecular biology into the mix. After deciding against fruitflies (not enough neuroanatomy was known) and mice (the tools for genetic manipulation were still being developed), she settled on *C. elegans*.

The nematode had been earning a growing scientific constituency⁵ ever since the mid-1960s, when British biologist Sydney Brenner proposed it as a model system. The organism was sufficiently complex to share basic biological functions with more advanced organisms, but was also experimentally tractable, with just 959 cells in the hermaphrodite marbled along its 1-millimetre length. Work by Brenner and others has established that all of the nematode's fundamental behaviours — navigating, foraging and mating — are governed by a neural system that contains just 302 neurons in hermaphrodites and about 8,000 synaptic connections. (By contrast, the human nervous system has 80 billion to 100 billion neurons and perhaps 100 trillion synapses.) To Bargmann, the combination of a blueprint for an entire nervous system and the powerful new genetic techniques that she had just learned offered "alluring" possibilities to map molecules related to behaviour onto a well-established neuroanatomy. She opted to stay at MIT and do a postdoc in the laboratory of H. Robert Horvitz, a leading *C. elegans* researcher.

Despite their relative simplicity, nematodes posed a daunting

challenge to anyone interested in dissecting complex behaviours: it wasn't clear that they actually had any. Most behavioural studies in *C. elegans* had investigated very simple responses, or reflex actions. But Bargmann, reasoning that nematodes need a way to find food (usually bacteria grazing in foul-smelling, decomposing matter), decided to tackle their sensory behaviour: specifically, how they detect attractive or noxious chemicals around them, process that information and then use it to navigate towards or away from the source.

Bargmann first had to work through "some rather awful smelling substances", recalls Horvitz, to figure out which ones worms prefer. Then she used a technique called laser ablation to obliterate individual nerve cells. She identified a series of neurons that the soil-dwelling *C. elegans* uses to detect chemicals in its immediate environment⁶, and then went on to find neurons that responded to volatile odours — in effect demonstrating for the first time that nematodes had a sense of smell⁷. Bargmann's systematic approach was "inspirational" to other worm biologists, Sternberg says.

Bargmann continued to explore worm olfaction after joining the faculty of the University of California, San Francisco, in 1991. In one study, her lab identified a receptor molecule, *odr-10*, in a pair of sensory neurons that detect diacetyl, an attractive odour associated with decaying food⁸. The lab correlated this and other mechanisms for sensing chemicals and heat with distinct nematode movements such as "pirouettes" and "omega turns", gradually assembling the neural circuitry of nematode navigation and exploration⁹ — a behaviour she sometimes calls, in casual conversation, "curiosity".

Soon, the lab began to uncover neural mechanisms that paralleled complex behaviour in other organisms. Bargmann knew that normal strains of roundworm vary in their feeding behaviour; some are solitary eaters, whereas others forage together in clumps of up to several hundred worms. Bargmann's lab showed that solitary eaters could be trans-

formed into social eaters by inserting a slightly different version of the *npr-1* gene, which in worms encodes a receptor with biochemical cousins throughout the animal kingdom, known as neuropeptide Y receptor¹⁰. In other animals, neuropeptide Y regulates food consumption, mood and anxiety, among other things.

In a sense, the *npr-1* story primed Bargmann for the nematocin discovery more than a decade later, which ultimately brought her back to the German ethologists — and to a new hypothesis about the evolution of behaviour.

DUE CREDIT

Bargmann's eighth-floor office in a research tower at Rockefeller is a homespun museum to the breadth of her intellectual interests and the depth of her personal attachments. A framed, red, heart-shaped piece of art on one wall depicts the anatomical outline of a nematode superimposed on the wiring pattern of the mouse olfactory system — a wedding gift celebrating her 2007 marriage to neuroscientist Richard Axel, whose research on mammalian olfaction at Columbia University in New York won Nobel recognition in 2004. A row of 21 empty champagne bottles lines the window sill, each uncorked to celebrate the thesis defence of a graduate student. Bargmann's speech oscillates between quiet scientific precision and a generous compulsion to acknowledge the contributions of every student, colleague, mentor and scientific ancestor, so that her conversation sometimes seems like an extended, erudite, magnanimous footnote.

One such conversation leads to the tale of the sexually confused males. Molecules related to vasopressin and oxytocin had previously been identified in other branches of the animal kingdom that have very deep evolutionary roots, including octopuses and annelids, suggesting that the molecules had an ancient and conserved role in animal behaviour. Around 2004, Evan Macosko, a PhD student in Bargmann's

"What are the most basic behaviours that every animal has to solve? Hunger, fear and reproduction."

lab, began scouring the *C. elegans* genome for the nematode version of oxytocin, but neuropeptides are very short and their genes are often hard to identify. He finally found a promising candidate in 2005, and the group went on to identify two nematocin receptors that were clearly related to the mammalian oxytocin receptor. Garrison's subsequent experiments with nematocin mutants dramatically confirmed that the peptide drives a basic behaviour¹.

Other research groups had defined, in exquisite detail, the series of discrete behavioural steps that male worms have to complete to succeed in mating (searching for a mate, contact, reverse turns, prodding for the vulva, insertion of spicule, transfer of sperm) as well as the motor neurons and muscles that rapidly fire and contract to drive these steps. But when Bargmann and her team analysed how the absence of nematocin affected each of these steps, they realized that each one remained intact. "It's not that he can't turn. It's not that he can't do the backing movement. It's not that he can't transfer sperm. It's that he doesn't know when to do them," she says. The neuropeptide, in essence, had a "global organizing role" and gave reproductive behaviour a forward drive. "There's something that's a much slower input that says something more like 'continue' or 'move forwards,' sort of providing momentum that's superimposed on it. So the nervous system is doing both fast and slow information processing, in parallel, to drive the behaviour."

This two-pronged neural processing reminded Bargmann of observations she had encountered decades earlier in the books in her mother's library. The pre-war ethologists — not only Lorenz and von Frisch, but also the Dutch scientist Niko Tinbergen — used astute field observations of fish, birds, insects and mammals to begin to assemble a fundamental grammar of behaviour. "[They] were the first to really express the idea that there were basic rules governing animal behaviours," she says, "and that you could recognize some of the same elements of those roles across very different animals."

A "fixed action response", for example, is a swift reaction to an environmental cue or threat. Tinbergen famously noted that the male stickleback fish in his lab aquaria flashed aggressive territorial behaviour when the local postal truck, painted red, rumbled by. He realized that the glimpse of red triggered a fixed, fast, hard-wired behavioural response because male sticklebacks display red bellies during mating season¹¹. At the same time, the ethologists described an "innate releasing mechanism", a slower orchestration of these fast responses that increases the probability that a fundamental behaviour such as mating will occur.

Tinbergen and his peers "were trying to relate what they saw in different animals to a common logic", Bargmann says, and she believes that logic is at least partly explained by neuropeptides such as nematocin. According to this model, the ritual movements of mating are the fixed-action patterns and the neuropeptides are the innate releasing mechanism. "Those [peptides], in ways that at this point we're still trying to work out, change the properties of the neurons involved in those behaviours to help organize the outcome," she says.

Bargmann suspects that this broad picture of nervous-system organization sends a counter-intuitive message about the evolution of behaviour: that the sensory apparatus in each species is evolving rapidly and

is highly divergent, creating a different set of behavioural cues and responses for different animals, whereas the overarching behavioural coordination exerted by neuropeptides remains largely evolutionarily conserved. "This is not the way we [usually] think of things in neuroscience," says Bargmann. "We always think the simplest part will be the sensory part, and maybe that will be the most conserved part. But in fact the sensory periphery is crazy unconserved between different animals."

This picture assumes that nematocin and its cousins have similar roles across species, but some biologists are not yet convinced. "It is pleasing to find these evolutionary connections," says Sternberg, "but we need more to know if the hypothesis is true."

The idea that peptides can have a global influence on what neurons in a network do is central to a debate roiling in the neuroscience community at the moment. A number of biologists are pushing to create the 'connectome' — a definitive wiring diagram that would map all the cells, synapses and neuromodulators in mammals and other complex organisms, and describe how those components interact to produce behaviour. It would be a massive, expensive undertaking.

Bargmann is all for connectivity maps, but she is not sure that, with current knowledge, they would explain as much as proponents hope. In an essay published last March¹², she warned that "it will not be possible to read a wiring diagram as if it were a set of instructions". Because neuropeptides can alter the excitability of neurons, the

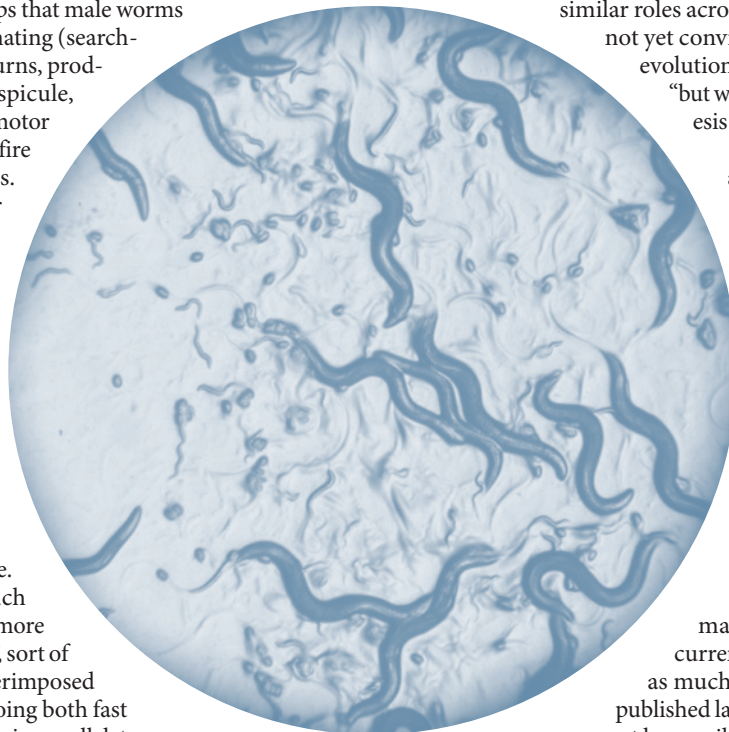
strength of synapses and even the overall function of a circuit, having a connectome is like having a street map without knowing how traffic flows through it, Bargmann says.

With nematocin in hand, the Bargmann lab is now trying to figure out what triggers its release, and how the peptide changes the activity of the neurons it targets. The team is also using nematocin to try to understand how social behaviours can differ within and between species. "Mating behaviours change rapidly over evolution compared to other behaviours," Bargmann notes. "How does that happen at a mechanistic level?"

Understanding the mechanics of the roundworm's simple nervous system, says Bargmann, "may be the only chance we have of figuring out a more complex system. I'm open to the possibility that the logic is different in other animals. I just see no evidence that, at a deep level, it's true." ■

Stephen S. Hall is a science writer in New York who also teaches public communication to graduate students in science at New York University.

1. Garrison, J. L. *et al. Science* **338**, 540–543 (2012).
2. Lorenz, K. *So kam der Mensch auf den Hund (Man Meets Dog)* (Deutscher Taschenbuch, 1950).
3. von Frisch, K. *Du und das Leben (You and Life: A Biology for Everyman)* (Ullstein Bucher, 1936).
4. Bargmann, C. I., Hung, M.-C. & Weinberg, R. A. *Nature* **319**, 226–230 (1986).
5. Ankeny, R. A. *Nature Rev. Genet.* **2**, 474–479 (2001).
6. Bargmann, C. I. & Horvitz, H. R. *Neuron* **7**, 729–742 (1991).
7. Bargmann, C. I., Hartwig, E. & Horvitz, H. R. *Cell* **74**, 515–527 (1993).
8. Sengupta, P., Chou, J. H. & Bargmann, C. I. *Cell* **84**, 899–909 (1996).
9. Gray, J. M., Hill, J. J. & Bargmann, C. I. *Proc. Natl Acad. Sci.* **102**, 3184–3191 (2005).
10. de Bono, M. & Bargmann, C. I. *Cell* **94**, 679–689 (1998).
11. Tinbergen, N. *The Study of Instinct* (Clarendon, 1951).
12. Bargmann, C. I. *Bioessays* **34**, 458–465 (2012).



C. elegans: 959-cell models of behaviour.

CONCRETE SOLUTIONS

Cement manufacturing is a major source of greenhouse gases. But cutting emissions means mastering one of the most complex materials known.

BY IVAN AMATO

If this year's expected global output of cement were somehow poured across Manhattan island, the 3.4-billion-tonne mass would solidify into a monolith about 14 metres high. If the monolith were created next year, it would probably be even bigger, given the construction boom now under way in developing nations such as China and India. Cement is a crucial raw material for civilization, holding together artefacts ranging from the 2,000-year-old Pantheon in Rome to modern skyscrapers and highways.

Unfortunately for Earth's climate, however, the most widely used form of that material today — 'portland' cement — is made by roasting limestone and clay in giant kilns in a process that sends nearly a tonne of carbon dioxide skywards for every tonne of final product. The manufacture of portland cement accounts for roughly 5% of all human-generated greenhouse-gas emissions.

Worse, for researchers looking for ways to reduce emissions, cement is not just a commonplace, high-volume commodity; it is also one of the most complex substances known in materials science. From its structure and composition to the reactions that ensue when it is mixed with water and poured into a mould to set, "we still have some of the most basic questions about cement unanswered", says Hamlin Jennings, director of the Concrete Sustainability Hub (CSHub) at the Massachusetts Institute of Technology in Cambridge.

"The details of what is happening once water touches cement powder are a matter of lively debate," adds Kenneth Snyder, a cement expert at the US National Institute of Standards and Technology in Gaithersburg, Maryland. "There are almost religious wars over this."

Nonetheless, the prospect of carbon taxes and cap-and-trade markets has led industry groups around the world to adopt green or

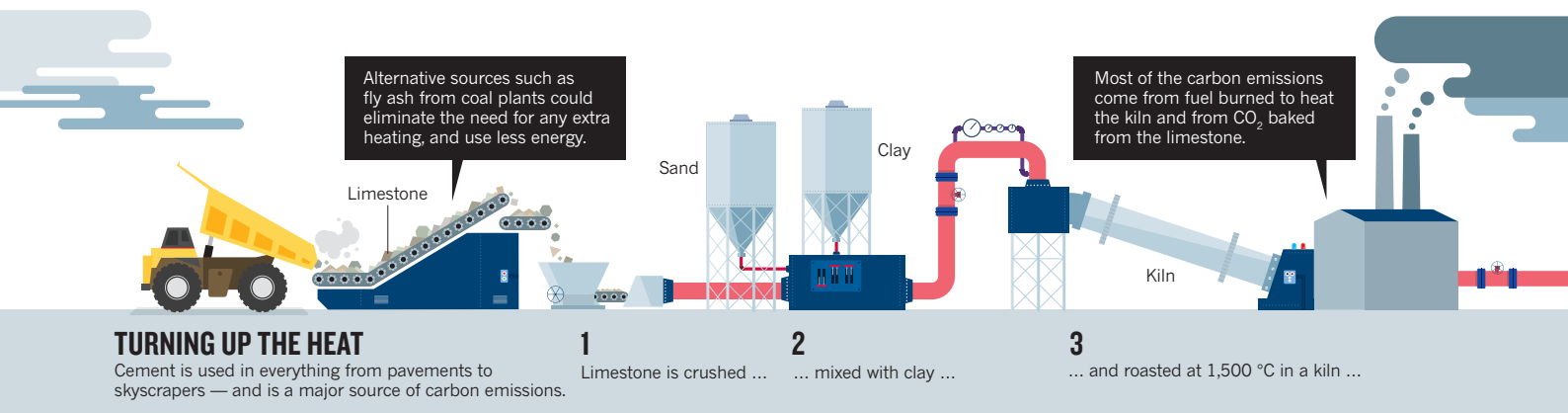
sustainable cement initiatives. Their approaches range from supporting basic research to pushing to reform international building codes, and, if successful, could eventually cut the cement industry's carbon dioxide footprint by half.

The CSHub is one of the field's largest academic-research centres. Founded in 2009 with funds from industry sponsors totalling US\$10 million over five years, the CSHub now comprises about a dozen principal investigators seeking to understand cement — from its function in various structures to its quantum-mechanical properties. It is a struggle, says Jennings, for reasons that become apparent when one considers what happens on the molecular scale when cement is made.

STRANGE BREW

The cement-making process (see 'Turning up the heat') begins with a mix of limestone and aluminosilicate clay, he says, "each with its own chemistry and impurities", which react in myriad ways as they are roasted together in a kiln at about 1,500 °C. What emerges are greyish, marble-sized chunks known as 'clinker'. Clinker contains silicon, iron and aluminium oxides (derived mostly from the clay) and calcium oxide, which forms when heat drives carbon dioxide out of the limestone's calcium carbonate. This carbon dioxide is one major source of emissions from the process; the fuel used for heating the kiln is the other. Once the clinker has cooled, it is combined with gypsum (the amount of which controls how fast the cement will set), milled into a powder having the consistency of flour and distributed to 'batch plants'.

There, cement powder is mixed with water to form a paste, the consistency of which depends on its intended use — in a bridge piling, say, or a pavement. Most often, the paste is mixed with sand, gravel or



larger stones to form concrete. The concrete slurry is then trucked to the construction site and poured into a mould, where it cures in a process that begins quickly but can take months to complete.

“One of the miracles, and the subject of intense research,” says Jennings, “is that the mix stays fluid for the first few hours, after which a furious set of simultaneous chemical reactions starts to produce the products that lead to the hardening process.” Most important to the final material are the hydration reactions that turn the water and powdered clinker into artificial stone: a matrix of calcium silicate hydrate ($\text{CaO-SiO}_2\text{-H}_2\text{O}$, or C-S-H). “All construction on this planet relies on this liquid-to-stone transition,” says Roland Pellenq, a physical chemist at the CSHub.

But C-S-H is a maddeningly imprecise formula, Pellenq says. Its components have no set proportions, and the reaction products in a given sample of curing concrete depend on the initial ingredients, the amount of water used and the ratio of calcium to silicon, as well as additives, contaminants, temperature and humidity. And, of course, concrete is opaque, adding to the difficulty of analysing C-S-H as it forms.

TWEAKING THE RECIPE

Despite these challenges, says Pellenq, he and his colleagues at the CSHub are making progress on the carbon-emissions problem. One promising line of attack involves finding ways to reduce the roasting temperature and thereby burn less fuel. The major targets are alite and belite, two of the primary minerals in clinker that give rise to C-S-H. Alite (Ca_3SiO_5) is the more reactive of the two — it begins to cure within hours after the addition of water, giving concrete its initial strength. But alite requires the full $1,500^\circ\text{C}$ to form, whereas belite (Ca_2SiO_4) forms at about $1,200^\circ\text{C}$. Belite is ultimately stronger, but takes days and even months to begin hardening — too long to be used on its own in construction projects. Pellenq and his colleagues are investigating whether some belite crystal structures might be as reactive as alite yet still form at lower kiln temperatures, saving fuel.

Because the answer to that question depends on atomic-scale details such as the distribution of electrons in the crystal, the researchers carried out quantum-mechanical calculations of how the structure of C-S-H is affected by aluminium, magnesium and other impurities (K. Van Vliet *et al.* *MRS Bull.* **37**, 395–402; 2012). As Pellenq puts it, “to do quantum clinker engineering, you need to know where the electrons are”. The CSHub researchers found that alite crystals always contain one plane that dissolves more easily in water than the others, whereas in belite crystals, all faces are similar — and the crystal is less reactive with water (E. Durgun *et al.* *Chem. Mater.* **24**, 1262–1267; 2012). This is why belite is slower to cure than alite. But Pellenq says that the results also suggest that certain impurities, such as magnesium, could help to make belite more soluble in water. This might allow it to cure fast enough to be used as the primary ingredient in cement for construction.

A move towards low-temperature belite could raise new problems, however. Franz-Josef Ulm, a mechanical engineer at the CSHub, and

his team have found that it takes four to nine times more energy to grind belite into powder than alite, which could diminish the emissions benefits of using belite-rich clinker.

Others, such as Ceratech, a cement company in Alexandria, Virginia, are seeking solutions in alternatives to conventional clinker. The company has found inspiration in the cement used 2,000 years ago by the engineers of ancient Rome. The key ingredient was pozzolana, a type of volcanic ash that reacts with water to make cement — thus functioning as natural clinker. Ceratech is exploiting an industrial version of pozzolana: fly ash, the fine particles filtered out of the combustion gases from coal-burning electricity plants. US plants produce roughly 70 million tonnes of fly ash every year, most of which is stored or disposed of in landfills. Ceratech converts the ash into cement powder by combining it with several proprietary liquid additives. Because the process does not require heat, the company says that its fly-ash cements are carbon neutral.

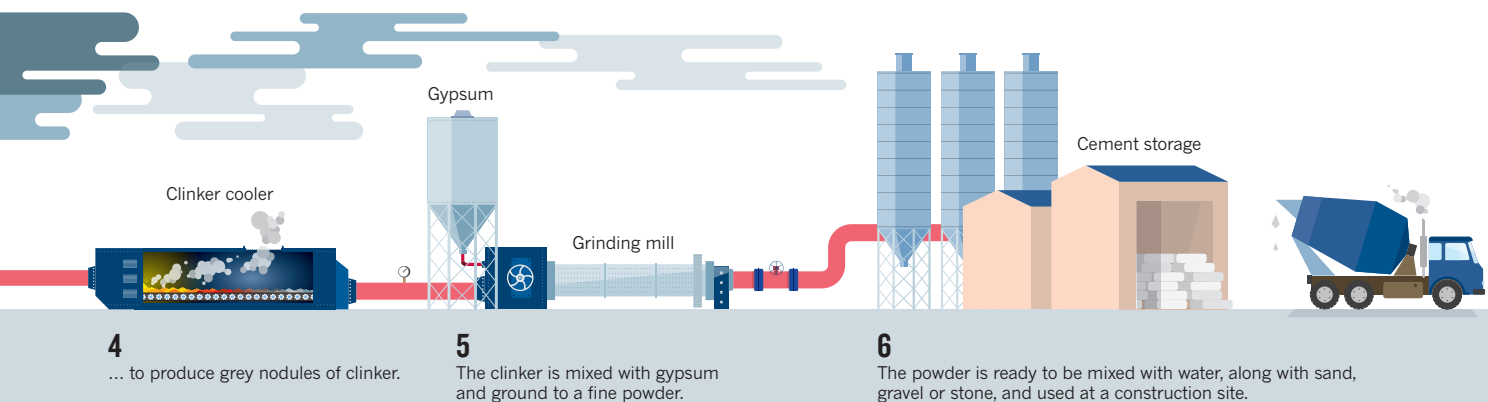
Although batch plants have been making blends containing up to 15% fly ash for years, Ceratech's formulations are 95% fly ash and 5% liquid ingredients, says Mark Wasilko, the company's executive vice-president. In addition, he says, concrete made from fly-ash cement is stronger than the conventional variety, so designers can use less of it. The company says that in a typical three-storey, 4,600-square-metre building, the use of its fly-ash cement would reduce the total volume of concrete by 183 cubic metres and the total mass of steel reinforcing bars by about 34 tonnes; it would also divert 374 tonnes of fly ash from landfills and reduce carbon dioxide emissions by 320 tonnes.

At present, Wasilko says, Ceratech is a bit player in the cement industry, and its approach to cutting carbon emissions amounts to a mere clink in a multibillion-tonne batch. The big carbon reductions will come only when next-generation cements are embraced by the construction industry's thousands of independent producers, engineers, architects, city planners and building inspectors. And that means lowering the perceived risk of choosing greener cements over their time-tested, conventional counterparts. The concern seems to be: “If this road doesn't work, my boss will hammer me,” says Snyder.

That attitude might change if more countries adopt taxes or cap-and-trade schemes that make the cost of emitting carbon much higher than it is now. But a more tangible, near-term way to overcome the reluctance is to build demonstration structures, such as bridges, roads and buildings, to prove the real-world viability of the new cement and concrete materials. Wasilko says that he hopes the several dozen projects that the company works on each year, such as dock structures in the Port of Savannah, Georgia, and chemical-handling basins for Gulf Sulphur Services in Galveston, Texas, will serve that purpose.

There is reason enough to get going. In the eight minutes or so that it took to read this article, cement-makers dumped another 30,000 tonnes of carbon dioxide into the atmosphere. ■

Ivan Amato is a freelance writer in Silver Spring, Maryland.



COMMENT

ENERGY Shale gas and oil will run out faster than champions predict **p.307**

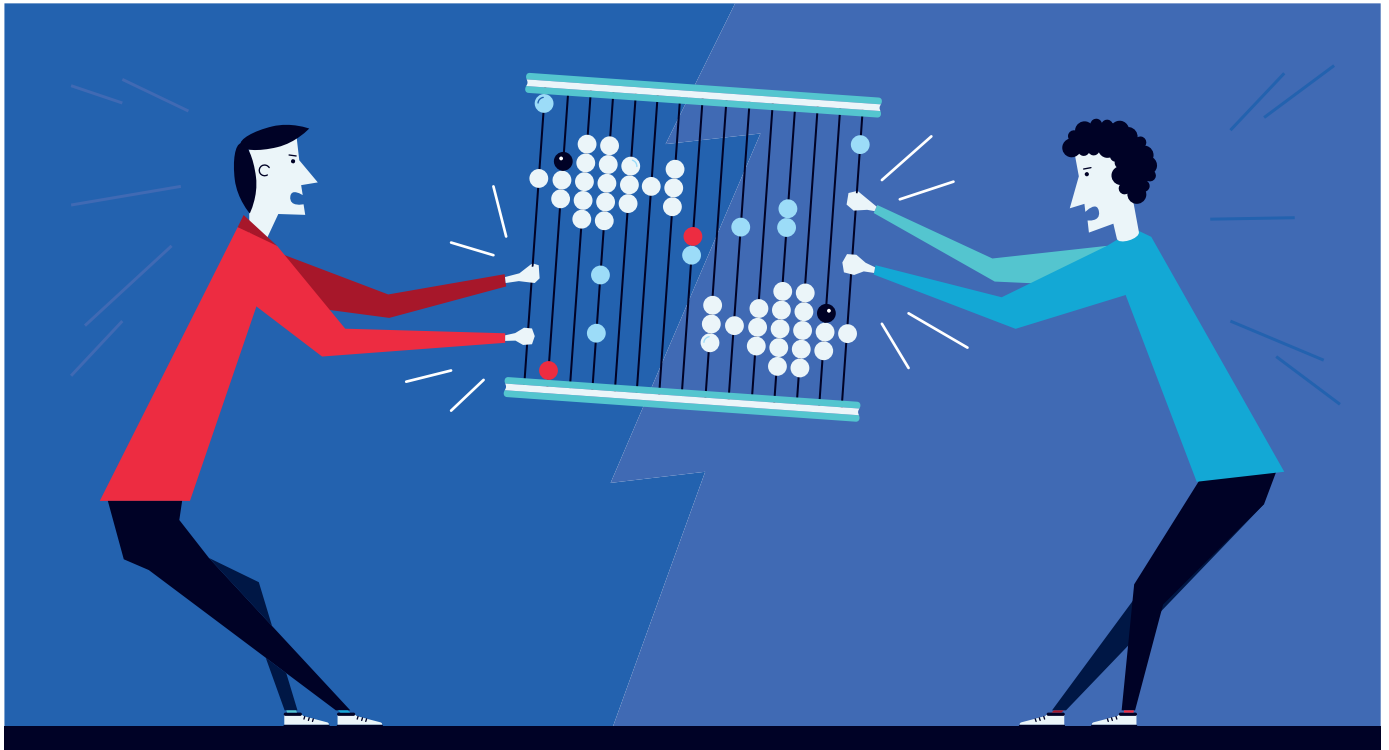
ANTHROPOLOGY Napoleon Chagnon sets the record straight **p.310**

BIOGRAPHY A flattering life of impish physicist Freeman Dyson **p.311**



HEALTH Iran needs international help to fight HIV epidemic **p.314**

DENIS CARRIER



Does catch reflect abundance?

Researchers are divided over the wisdom of using estimates of the amount of fish hauled in each year to assess the health of fisheries.

POINT

Yes, it is a crucial signal

The only data available for most fisheries are the weight of fish caught each year, insists Daniel Pauly.

In developed countries such as the United States, Australia and members of the European Union, many fisheries are monitored by fisheries scientists using expensive stock assessments. To infer the size of the fish populations being exploited, scientists use the age and size distributions of the fish caught; the results of scientific surveys carried out from research vessels; and information about growth and migration from tag and recapture studies. Yet the only data ►

COUNTERPOINT

No, it is misleading

Many factors as well as abundance determine the hauls of fishermen, warn Ray Hilborn and Trevor A. Branch.

The major database on all the fisheries of the world is the *FAO Yearbook. Fishery and Aquaculture Statistics*. This collates the amount (in weight) of haddock, bream, cod and more than 1,000 other species hauled in each year by fishermen, whether from commercial trawlers or canoes, using estimates sent in by officials from individual countries.

For the past few years, researchers have been conducting analyses ►

POINT: YES, IT IS A CRUCIAL SIGNAL ▶ that are collected and made publicly available for the fisheries in about 80% of all maritime countries are estimates of the weight of fish caught each year. Since 1950, the Food and Agriculture Organization (FAO) of the United Nations has published these catch data (which are gathered by officials in around 200 countries) in the *FAO Yearbook. Fishery and Aquaculture Statistics*.

A debate is raging among fisheries scientists over the wisdom of using catch data to assess the health of fisheries. I agree that catch data should be used with care. But the current dispute is sending a message to policy-makers that catch data are of limited use. If countries — especially developing ones — start to devote even fewer resources to collating catch data, our understanding of fisheries, including their impact on marine ecosystems and their importance for local economies, will suffer.

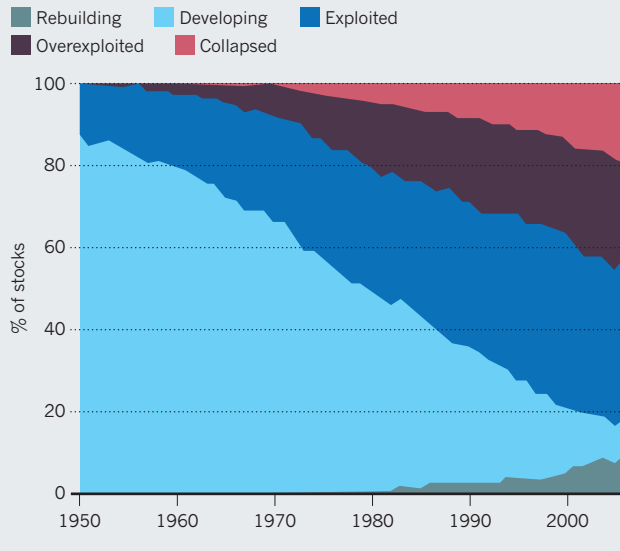
STOCK PLOT

The debate over catch data stems from an analytical approach that was pioneered by the FAO, and subsequently developed by others, myself included. In 1996, FAO researchers devised what became known as a stock status plot¹. For 400 well-studied fisheries, the researchers plotted catch data over time, and used the slope of the graphs to assign the stocks to different categories, such as ‘developing’ where catches were increasing, or ‘senescent’ where catches had collapsed. The resulting chart was meant to show at a glance how the fisheries had fared since the 1950s (badly, apparently).

In 2001, the FAO method was modified by fisheries scientist Rainer Froese at the GEOMAR Helmholtz Centre for Ocean Research in Kiel, Germany. He and I then used the modified method to produce stock status plots for all the fisheries in the world for which catch data were available (see ‘The stock status plot’; and go.nature.com/blfbma). Our results revealed a similar trend to that shown by the FAO: the number of collapsed stocks had steadily increased over the years, and by the mid-1990s, 20% of the stocks exploited in the 1950s had collapsed. (We

THE STOCK STATUS PLOT

The Food and Agriculture Organization of the United Nations pioneered a way to visualize trends in fisheries using catch data.



ADAPTED FROM D. PAULY AMBIO 34, 290–295 (2007)

classified stocks as collapsed if their annual catch had fallen to less than 10% of the highest ever recorded.) Unfortunately, it took another ten years, and a misguided claim, for the world to take notice.

In 2006, a group of researchers from various institutions used a stock status plot to project, among other things, that all stocks would be collapsed by 2048 (ref. 2). Unsurprisingly, this projection, although a small part of the study, triggered an avalanche of alarmist headlines: “Seafood may be gone by 2048,” wrote the *National Geographic*; “The end of fish, in one chart,” said *The Washington Post*.



JEFFREY L. ROTMAN/GETTY

Fishermen have pulled fewer fish from the world's oceans in recent years.

The weirdly precise 2048 date, with echoes of George Orwell's *Nineteen Eighty-Four*, was widely derided within the fisheries community. Given the myriad factors that can affect fishing — shifts in policy, rising fuel costs, market crashes and natural disasters — it is impossible to predict where fisheries will be even ten years from now. But, of the various lines of attack that fisheries scientists have used to discredit the 2006 paper, one charge has since gained momentum and stands to do much more damage to fisheries science and management than the original paper. This is the idea that catch data are not useful for determining the health of fish stocks. This is wrong. Dangerously so.

WEIGHT OF EVIDENCE

Over the past two decades, the amount of fish caught from the world's oceans has declined. Factions of the fisheries community disagree over how to interpret this decline, and they dispute the methods used to assign stocks to different categories, such as collapsed or under-exploited. And it is true that catch size is not just affected by fish abundance — numerous factors, such as a change in management or legislation, can also influence the annual haul of fish. But for the vast majority of species, no signal of this downward trend would even exist without the FAO catch data.

When only catch data are available, fisheries researchers can and should use these data to infer fishery status, at least tentatively^{3,4}. Even when stock assessments or scientific surveys are conducted, such information should always be used in conjunction with any and all available catch data. Take, for example, the Canadian northern cod stock, which unexpectedly collapsed in Newfoundland and Labrador in the early 1990s, even as stock-assessment experts were monitoring it using state-of-the-art methods to model abundance⁵. In the years before the collapse, fishermen were using either net traps fastened on the sea floor or trawlers, but because the boats could track the shrinking shoal, their catch remained high, even as the trap fishermen started bringing in fewer and fewer cod. Stock-assessment experts had monitored only the trawler catches.

Discrediting catch data risks hampering analysis and might also discourage efforts to improve the quality of fisheries statistics worldwide. For the vast majority of species, expert stock assessments can cost from around US\$50,000 to millions of dollars per stock — especially when research vessels are involved — so are often not feasible. If resource-starved governments in developing countries come to think that catch data are of limited use, the world will not see more stock assessments; catch data will just stop being collected.

Instead of questioning the usefulness of catch data in assessing stocks, scientists should be urging more governments to collect them (along with data on fishing effort, the economic value of catches and fishing costs), and devising cost-effective ways to improve their reliability.

As part of the Sea Around Us initiative — a collaboration between the University of British Columbia in Vancouver, Canada, and the Pew Charitable Trusts, which aims to monitor the impact of fisheries on marine ecosystems — I am leading a project to evaluate the entire body of FAO catch data collected since 1950. So far, my team has gathered information on fish consumption and the tonnage of fish imported and exported, for instance, to verify the catch data of 180 countries and island territories. Our findings suggest that catches, with the notable exception of domestic catches by China, are under-reported by about 100–500% in many developing countries⁶, and by 30–50% in developed ones⁷.

While fisheries researchers continue the important debate about which fisheries are in decline, why and to what degree, most fishermen worldwide are finding fewer fish in their hauls than their predecessors did. Knowing what tonnage is pulled out of the oceans each year is crucial to knowing how to reverse this trend. ■

Daniel Pauly is at the Fisheries Centre of the University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.
e-mail: d.pauly@fisheries.ubc.ca

COUNTERPOINT: NO, IT IS MISLEADING ► to use these data to assess the health of the world's fish stocks. And high-impact journals, including this one, have published them. Such assessments consistently overlook the fact that the amount of fish caught does not necessarily reflect the number of fish in the sea.

Attempts to use catch data as an indication of fish abundance have spread alarm and confusion in policy circles, and fuelled the perception among the public and conservation organizations that fisheries management is failing. A much better approach is to deduce the health of stocks region by region and fishery by fishery using scientific stock assessments, which collate all sorts of data — from the results of surveys conducted from research vessels to the catch per fishing effort, and the age and size distributions of the fish caught. This can reveal which management strategies are actually effective.

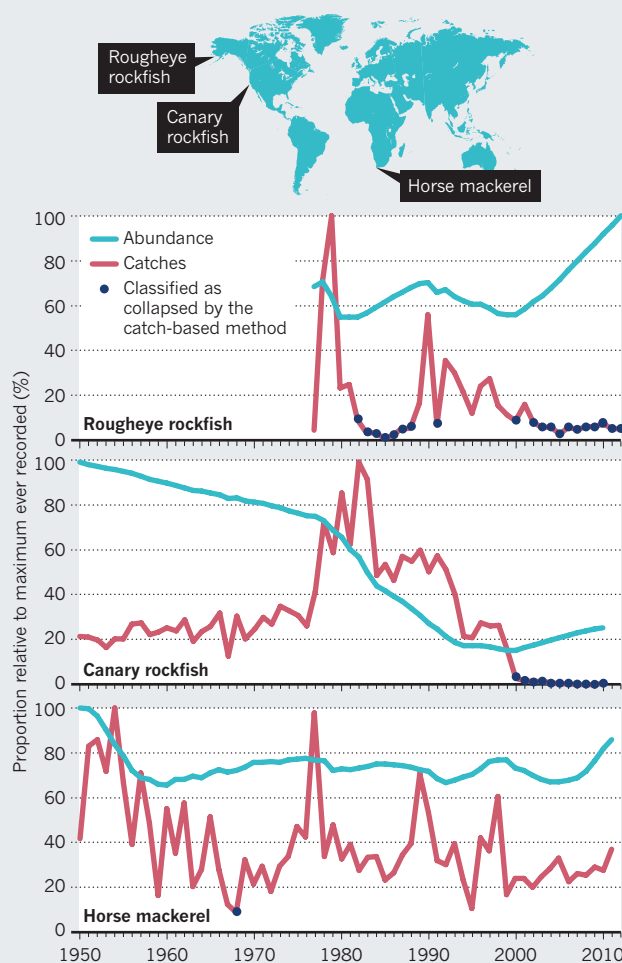
ON THE BRINK?

In 2006, researchers projected that all the fisheries in the world would collapse by 2048 (ref. 2). The group classed a stock as 'collapsed' if less than 10% of the highest catches ever recorded were being harvested each year⁸, using an approach developed by others. For many of us in the fisheries community, the 2048 projection was hard to believe; Alaskan salmon fisheries, for instance, were in better shape than they had been in the history of industrial fishing.

Another paper soon followed⁸, written by the authors of the original study² and some of their critics, ourselves included. This suggested ►

WHAT'S THE CATCH?

The tonnage of fish caught each year can soar or plummet, regardless of how many fish are in the sea. For rougheye rockfish and canary rockfish, fishing regulations have helped to reduce catches in recent years.



COUNTERPOINT: NO, IT IS MISLEADING ▶ that many of the world's fisheries are not in nearly such a perilous state as the 2048 projection implied. In this study, we analysed trends only in those stocks that are closely monitored by research agencies⁸. Using stock-assessment data, we found that over the previous two decades, the number of stocks that had collapsed had indeed risen, but by the mid-2000s only 14% of the 166 stocks we analysed were collapsed. Also, for two-thirds of these stocks, which were mainly in Europe, North America and Australasia, populations had stabilized owing to fishing restrictions, and some were beginning to rebuild.

The 2048 projection and the controversy it sparked within the fisheries community undermined conservationists' confidence in the ability of governments to prevent overfishing. Even more misleading, in our view, is the most recent attempt to assess all the fish stocks in the world using catch data.

The Ocean Health Index, published in *Nature* last year⁹, aggregates various measures, such as biodiversity, into a single score that is meant to indicate how healthy the seas are. To calculate the capacity of the oceans to generate food, the researchers estimated the maximum sustainable catches of 100 well-studied stocks using the results of stock assessments. They then used a previously devised equation relating these maximum sustainable catches to the highest catches ever recorded from these same fisheries to estimate the maximum sustainable catches for all the fisheries in the FAO database.

Hugely unreliable predictions result. For instance, the maximum sustainable yield for Antarctic krill assigned by the Ocean Health Index is 174 times lower than the estimate derived from detailed stock assessments conducted by the International Commission for the Conservation of Antarctic Marine Living Resources¹⁰. Overall, the health-index study implies that fishermen could theoretically catch several times more marine fish if fisheries were better managed. This wildly contradicts the findings of several other studies — that fishermen are already catching 80–96% of the potential global yield^{11,12}.

STOCK OPTIONS

So why do analyses using catch data as a proxy for fish abundance generate results that other studies throw into question?

The main difficulty is that a low catch compared with previous records does not necessarily mean fewer fish, and vice versa. For instance, 34 large stocks along the west coast of Oregon and Washington in the United States would seem to have collapsed based on recent catch data¹³. Yet according to detailed stock assessments, only three (anchovies, eulachons or candlefish, and abalone) have actually collapsed¹³.

Catches can shift hugely for many reasons. For example, hauls of stocks along the west coast of the United States have tended to shrink because of markets or new fishing regulations (see 'What's the catch?'). Changes in taxonomy are another reason. In the 1950s, global shark catches were assigned to seven taxonomic groups. By the early 2000s, 36 groups were recognized in the FAO catch database. Catches of sharks categorized in the early groupings may have dwindled not because fewer sharks are being caught, but because they are being registered under different names.

Catch data can also be affected by alterations to national jurisdictions, such as the expansion of most nations' fishing rights to 200-mile expanses of ocean in the 1970s and 1980s; natural disasters such as the *Exxon Valdez* oil spill, which resulted in the closure of many Alaskan salmon fisheries for a year; civil war; increases in fuel costs; and low fish prices. Regime changes are also a factor. After the disintegration of the Soviet Union in 1991, the loss of government support meant that thousands of fishing vessels could no longer operate.

Furthermore, few countries monitor anything but their largest or economically most important fisheries with the rigour used in the

United States and within the European Union. In most cases, officials simply give their best guess. Even in the United States, where the federal government allocates US\$880 million to fisheries agencies each year, reliable catch data for smaller stocks, such as spiny dogfish, are hard to obtain.

BETTER MANAGEMENT

Analyses based largely on catch data are fuelling a view held by numerous non-governmental organizations and environmentalists that the only solution to marine conservation is a ban on fishing in large areas. Between 2007 and 2009, for instance, several non-governmental organizations and US foundations spent \$58 million each year on efforts to create 'marine protected areas'. This campaign has been spectacularly successful in Australia, where 3.1 million square kilometres of ocean are now closed to fishing.

This approach overlooks the enormous successes of many management strategies. For example, on the east coast of the United States the total abundance of bottom fish, such as redfish and haddock, increased

more than fivefold from 1995 to 2007 after fishing restrictions were strengthened from the mid-1990s.

If the FAO catch data were the only source of fisheries data, fisheries science would be flying blind. But scientific assessments of trends in fish abundance spanning three or four decades are now publicly available for the fisheries that constitute 40% of the total catch in the FAO global database. These data come mostly from developed countries (in North America, Europe and Australasia) or from major international fisheries such as those for tuna.

Various data collected by national fisheries agencies, fishing companies and marine scientists are also available for the rest of the world. Currently, there is no global database for this information. We (the authors) have recently begun working with more than 20 countries, the FAO and the World Bank to try to assemble a database for a good sample of the world's fisheries. We estimate that obtaining the kind of data needed for 40 countries (focusing on 6–8 fisheries in each country) will take 10 years and require US\$20 million.

Meanwhile, we urge researchers to use all the available data in addition to the FAO database, and to validate their results by consulting local experts or other data sources. Catch data are a crucial part of any fisheries assessment — it is impossible to calculate the maximum weight of fish that could be harvested sustainably without knowing what is being caught each year. But on their own, catch data cannot answer the question at the heart of fisheries science: how many fish are in the sea? ■

Ray Hilborn and Trevor A. Branch are at the School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington 98195, USA.

e-mails: rayh@uw.edu; tbranch@uw.edu

1. Grainger, R. J. R. & Garcia, S. M. *Chronicles of Marine Fishery Landings (1950–1994): Trend Analysis and Fisheries Potential* (FAO, 1996).
2. Worm, B. *et al. Science* **314**, 787–790 (2006).
3. Kleisner, K., Froese, R., Zeller, D. & Pauly, D. *Fish Fish*. <http://dx.doi.org/10.1111/j.1467-2979.2012.00469.x> (2012).
4. Froese, R., Kleisner, K., Zeller, D. & Pauly, D. *Marine Biol.* **159**, 1283–1292 (2012).
5. Walters, C. J. & Maguire, J.-J. *Rev. Fish Biol. Fish.* **6**, 125–137 (1996).
6. Zeller, D., Booth, S., Davis, G. & Pauly, D. *US Fish. Bull.* **105**, 266–277 (2007).
7. Zeller, D. *et al. Fisheries Res.* **108**, 356–363 (2011).
8. Worm, B. *et al. Science* **325**, 578–585 (2009).
9. Halpern, B. S. *et al. Nature* **488**, 615–620 (2012).
10. Nicol, S., Foster, J. & Kawaguchi, S. *Fish Fish*. **13**, 30–40 (2012).
11. Sumaila, U. R. *et al. PLoS ONE* **7**, e40542 (2012).
12. Pauly, D. *Fish Res.* **25**, 25–38 (1996).
13. Branch, T. A., Jensen, O. P., Ricard, D., Ye, Y. & Hilborn, R. *Conserv. Biol.* **25**, 777–786 (2011).



Gas being burnt off at the Bakken shale oil field in North Dakota as a by-product of oil extraction.

A reality check on the shale revolution

The production of shale gas and oil in the United States is overhyped and the costs are underestimated, says **J. David Hughes**.

The 'shale revolution' — the extraction of gas and oil from previously inaccessible reservoirs — has been declared an energy game changer. It is offsetting declines in conventional oil and gas production, with shale gas being heralded as a transition fuel to a low-carbon future, and shale oil as being capable of reinstating the United States as the largest oil producer in the world, eliminating the need for foreign imports.

These heady claims have been largely accepted by government forecasters, including the International Energy Agency¹ and the US Energy Information Administration (EIA). The oil firm BP predicts that production of shale gas will treble and shale oil — also known as 'tight oil' — will grow sixfold from 2011 levels by 2030 (ref. 2).

The claims do not stand up to scrutiny. In a report published this week by the Post Carbon Institute³ in Santa Rosa, California, I analyse 30 shale-gas and 21 tight-oil fields (or 'plays') in the United States, and reveal that the shale revolution will be hard to sustain. The study is based on data for 65,000 shale wells from a production database that is widely used in industry and government. It shows that well and field productivities exhibit steep declines. Production costs in many shale-gas plays exceed current gas prices, and maintaining production requires ever-increasing drilling and the capital input to support it.

Although the extraction of shale gas and tight oil will continue for a long time at some level, production is likely to be below the exuberant forecasts from industry and government. I see supplies of shale gas declining substantially in the next decade unless prices rise considerably. A more realistic debate around shale gas and tight oil is urgently needed — one that accounts for the fundamentals of production in terms of sustainability, cost and environmental impact.

SHALE GAS

Two technologies — horizontal drilling coupled with large-scale, multi-stage hydraulic fracturing (fracking) — have made it possible to extract hydrocarbons trapped in impermeable rocks (see *Nature* 477, 271–275; 2011). In 2004, less than 10% of US wells were horizontal; today, the figure is 61%.

Most shale-gas production worldwide is in North America, although pilot projects are being conducted in many countries. Production has been on a plateau since early 2012 after a period of sharp growth. Shale gas has risen from about 2% of US gas production in 2000 to nearly 40% in 2012 (ref. 3); overall US gas production grew by 25% over the same period. The resulting supply glut drove US gas prices down severely. Prices have since recovered slightly but remain too low for many shale-gas plays without liquids production to be economically viable.

Large-scale shale-gas production was

initiated in the Barnett Shale formation a decade ago, and it spread quickly to other areas. Five plays produce 80% of US shale gas (listed from highest to lowest output): Haynesville in Louisiana, Barnett in east Texas, Marcellus (which spans West Virginia, Pennsylvania and New York), Fayetteville in Arkansas and Woodford in Oklahoma.

A pattern of events has emerged. When a play is discovered, a leasing frenzy ensues. This is followed by a drilling boom because the lease assignments, often 3–5 years long, can be terminated if the site is not producing gas. Sweet spots — small areas with high productivity — are identified and drilled off first, with marginal areas targeted next. Average well quality (as determined by initial productivity) rises at first and then declines.

In four of the top five shale-gas plays, average well productivity has been falling since 2010 (see ‘Top five shale plays’). In the Haynesville play, an average well delivered almost one-third less gas in 2012 than in 2010. The exception is the Marcellus: supply is rising in this young, large play as sweet spots are still being found and exploited.

Wells decline rapidly within a few years. Those in the top five US plays typically produced 80–95% less gas after three years. In my view, the industry practice of fitting hyperbolic curves to data on declining productivity, and inferring lifetimes of 40 years or more, is too optimistic. Existing production histories are a few years at best, and thus are insufficient to substantiate such long lifetimes for wells. Because production declines more steeply than these models typically suggest, the method often overestimates ultimate recoveries and economic performance (see go.nature.com/kiamlk). The US Geological Survey’s recovery estimates are less than half of those sometimes touted by industry⁴.

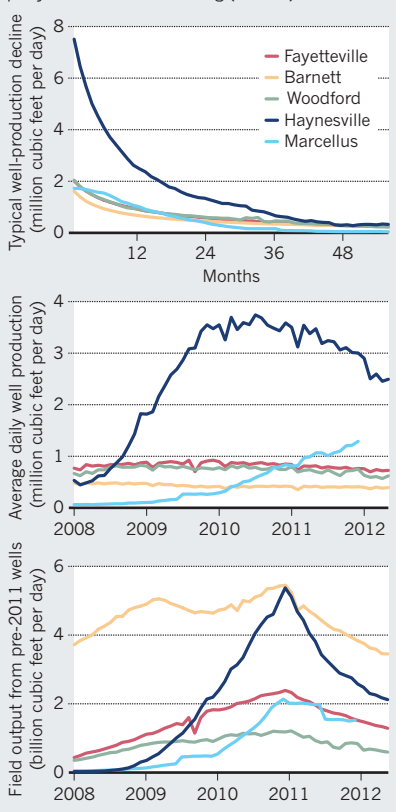
New wells must be drilled to maintain supply. In the Haynesville play, almost 800 wells — nearly one-third of those that were active in 2012 — must be added each year to keep shale-gas output at 2012 levels. With capital costs of around US\$9 million per well, drilling to keep production flat costs some \$7 billion a year. Full costs, including leasing, infrastructure and interest, are even higher³.

Across the United States, this equates to 7,200 wells at a cost of more than \$42 billion annually, simply to offset declines in production³. This investment by drilling companies — to sustain production to prop up share prices — is not covered by sales. In 2012, US shale gas generated just \$33 billion (although some wells also produced substantial liquid hydrocarbons, which improved economics). To break even in shale-gas plays without liquids production, gas prices would have to rise⁵.

Shale gas thus requires large amounts of capital from industry to maintain production⁶. Over time, the best shale plays and

TOP FIVE SHALE PLAYS

Five US fields produce 80% of shale gas. The output of a typical well drops 80–95% in its first three years (top). Average well output across a field peaks and then falls as prime spots are used up (middle). Total field production falls 30–50% per year without new drilling (bottom).



their sweet spots are drilled off, so the costs of keeping up supply will increase. Much of current shale-gas production is uneconomic, and will require higher gas prices just to maintain production, let alone increase it.

TIGHT OIL

The story is similar for tight oil. Two plays produce 81% of US tight oil — Eagle Ford in south Texas and the Bakken in North Dakota and Montana. The productivity of new wells in both areas drops by about 60% after one year, levelling out to less than 40% in the second year, less than 30% in the third year and so on. Overall field decline, which combines the productivity of older and newer wells, is about 40% per year³.

The ultimate output of these plays depends on the maximum number of available drilling locations. Wells cannot be drilled too close together because they drain the same reservoir volume, which increases costs and does not improve recovery. The EIA estimates that the Bakken and Eagle Ford plays can host almost three times the current number of wells, or nearly 12,000 each^{3,7}.

Assuming that the Bakken’s current drilling rate of 1,500 wells a year is maintained, my analysis predicts that its production could

rise to nearly 1 million barrels of oil a day. Given the EIA estimates of the maximum number of available drilling locations in the Bakken, however, I suggest that production will peak by 2017, when available well sites are exhausted, and then fall by 40% a year. I disagree with those who maintain that the Bakken’s production can stay at that high level for many years — this would require thousands more wells than would fit⁸.

LONG VIEW

Governments and industry must recognize that shale gas and oil are not cheap or inexhaustible: 70% of US shale gas comes from fields that are either flat or in decline. And the sustainability of tight-oil production over the long term is questionable.

High-productivity shale plays are not ubiquitous, as some would have us believe. Six out of 30 plays account for 88% of shale-gas production, and two out of 21 plays account for 81% of tight-oil production. Much of the oil and gas produced comes from relatively small sweet spots within the fields. Overall well quality will decline as sweet spots become saturated with wells, requiring an ever-increasing number of wells to sustain production.

Production will ultimately be limited by available drilling locations, and when they run out, production will fall at rates of 30–50% per year. This is projected to occur within 5 years for the Bakken and Eagle Ford tight-oil plays.

The EIA’s projections imply that, by 2040, the United States will recover all currently known shale-gas reserves, 58% of unproved shale-gas resources and 78% of unproved tight-oil resources^{7,9}. These predictions are wildly optimistic given the fundamentals of producing these hydrocarbons. Similarly, the EIA forecast of gas prices strains credibility⁹ because it is below many other estimates of the cost of production with steadily rising supply for the next two decades. Declaring US energy independence and laying plans to export the shale bounty is unwise. The long-term viability of the shale revolution must be accounted for in a sustainable energy strategy for the future. ■

J. David Hughes is a geoscientist and a fellow of the Post Carbon Institute, Santa Rosa, California 95404, USA.
e-mail: davehughes@twincomm.ca

1. World Energy Outlook 2012 (IEA, 2012).
2. BP Energy Outlook 2030 (BP, 2013).
3. Hughes, J. D. *Drill, Baby, Drill* (Post Carbon Institute, 2013).
4. US Geological Survey. Open-File Report 2012-1118 (USGS, 2012).
5. Jacoby, H. D. et al. *Econ. Energy Environ. Policy* **1**, 37–51 (2012).
6. Krauss, C. & Lipton, E. *New York Times* (20 October 2012).
7. US Energy Information Administration *Annual Energy Outlook 2012* (EIA, 2012).
8. Morse, E. L. *Energy 2020* (Citi, 2012).
9. US Energy Information Administration *Annual Energy Outlook 2013 Early Release* (EIA, 2012).



A Yanomami of the Majecodoteri tribe in the Amazon rainforest.

ANTHROPOLOGY

Tribal warfare

Douglas William Hume assesses a first-hand account of controversial work with the Yanomami people.

Napoleon Chagnon is perhaps best known for his classic 1968 Amazon ethnography *Yanomamö: The Fierce People*, now in its sixth edition. He is also among a handful of scholars whose research prompted the 2001 creation of an American Anthropological Association (AAA) investigatory body, the El Dorado Task Force, to investigate claims of ethical and scholarly wrongdoing during research with the indigenous Amazonian Yanomami.

Although a referendum “to rescind the AAA’s acceptance of the Report of the El Dorado Task Force” passed with a clear majority of the AAA’s membership in 2005, Chagnon’s work and character have, in my view, been scarred by the accusations of misconduct and even genocide that prompted the investigation and split anthropologists into camps representing the false dichotomy of ‘for science’ or ‘for indigenous rights’.

That controversy drives Chagnon’s autobiography, *Noble Savages*. The title is a satirical nod to the romanticization of indigenous peoples as intrinsically good and close to nature, a concept often attributed to eighteenth-century French philosopher Jean-Jacques Rousseau — and still accepted by some anthropologists. Chagnon makes it clear that he views neither the Yanomami nor anthropologists as “noble savages”.

Noble Savages: My Life Among Two Dangerous Tribes — the Yanomamö and the Anthropologists

NAPOLEON A. CHAGNON

Simon & Schuster: 2013. 544 pp. £23.20/\$32.50

The book begins with Chagnon’s arrival in 1964 at his first Yanomami village, Bisaasiteri in Venezuela, where he was greeted by drawn arrows. His description of how he experienced Yanomami life is candid and vivid, from challenges to his preconceptions about the daily lives of so-called primitive peoples to the joy of discovering symbolic meanings of ritualized behaviour. Chagnon goes on to describe how he learned about Yanomami genealogy, violence and other cultural traits by participating in village life and rituals, and collecting family and conflict histories, between 1964 and the late 1990s.

Chagnon’s central narrative focuses on how he applied biologist E. O. Wilson’s sociobiological theory to genealogy and violence among the Yanomami. Roughly, Wilson posited that evolutionary processes shape social interactions such as altruism. Chagnon explains how he collected and analysed ethnographic data by finding statistical patterns between kinship and violence; reached conclusions on probable causes of warfare such as abduction of women; developed new questions about the impacts of violence on

reproductive success; and repeated the process over decades of research.

Towards the end of the book, Chagnon addresses the debate over his work. He gives a full description of his contentious relationships with many anthropologists, missionaries and indigenous-rights activists.

His first conflict was with anthropologists who favour postmodernism (rejecting absolute truths and focusing instead on the experience of the anthropologist) and advocacy (actively engaging in human-rights campaigns with indigenous peoples), who viewed his work as anti-Yanomami, deterministic and racist. The second was with Catholic missionaries who saw Chagnon’s influence as harmful to the Yanomami, and successfully lobbied the Venezuelan government to expel him from the country. The third attack — with which I am most familiar, because I manage the online archive of material on it (<http://anthroniche.com>) — is related to the ‘Darkness in El Dorado’ scandal, sparked by a book of the same name by journalist Patrick Tierney (W. W. Norton, 2000).

Darkness in El Dorado levelled accusations that Chagnon, several medical researchers and other anthropologists working among the Yanomami were individually and sometimes collectively guilty of exploitation, paedophilia and genocide — the latter by the introduction of measles and other contagious diseases to the indigenous community. *Noble Savages* is, to a degree, the long-awaited response to Tierney’s book, and its final chapter relates Chagnon’s perspective on events that unfolded after *Darkness* was published, both in the task-force investigation and among anthropologists. It is Chagnon’s detailed accounts of his research, however, that serve to refute Tierney’s claims, which were based on interviews with missionaries, anthropologists and the Yanomami.

Sadly, however much evidence Chagnon presents, his detractors will remain his detractors: the controversy has become more about a moral cause (indigenous human rights) than about facts.

Noble Savages is the story of a man who for decades has tried to bring evolutionary theory and scientific methods to the study of humanity in anthropology. In short, it is Chagnon’s philosophy-of-science case study, as he struggles against anthropology’s retreat from science. His book is an important contribution to the debates over the methods and theories used to understand humans in anthropology and evolutionary sciences — and to debates over how visionaries become the targets of those who do not share their vision. ■

Douglas William Hume is an assistant professor of anthropology at Northern Kentucky University in Highland Heights, and maintains the website *AnthroNiche*. e-mail: humed1@nku.edu

PHYSICS

Rebel without a pause

Robert P. Crease delves into a life of Freeman Dyson, a theoretical physicist who chose a non-conformist path.

Freeman Dyson looks like a wizard. Tall and lanky, he has piercing blue eyes, elfin ears, an aquiline nose and a full head of hair, despite being in his 90th year. From the cover of Phillip Schewe's buoyant biography *Maverick Genius*, he stares at us with an expression that is half cherub, half imp.

Dyson's career is marked by equally sharp contrasts. He has made significant contributions to technical areas of mathematical physics and quantum electrodynamics — but also to speculative fields such as space travel and astrobiology. His writings mix science and poetry. He is not religious, yet won the Templeton Prize for “progress in religion”. He crusades against nuclear weapons, but engages in defence research. And he is known for taking heretical positions on subjects such as extrasensory perception and climate change. Yet his most eccentric facet is that he became an eminent theoretical physicist without having a PhD — a feat for which Richard Feynman, that other famous non-conformist of physics, much envied him.

“This is a biography,” Schewe alerts us, “and not science history.” Thus we learn of Dyson's beginnings in the United Kingdom, his early attraction to mathematics and his operations research at Royal Air Force Bomber Command during the Second World War. After the war, Dyson moved to the United States, where he worked first at Cornell University in Ithaca, New York, and then the Institute for Advanced Study in Princeton, New Jersey, along with the likes of Albert Einstein and J. Robert Oppenheimer. We hear of Dyson's work on the TRIGA project to build a foolproof nuclear reactor model, and on the Orion project to design a spacecraft that would boost itself to high speeds by surfing on atomic explosions — and that Dyson intended to be aboard the first flight. We read about his work on neutron bombs; his opposition to the Nuclear-Test-Ban Treaty (because it would inhibit defence research); the solar-energy collector known as the ‘Dyson sphere’ once mentioned on *Star Trek*; and his thought experiments classifying possible extrasolar civilizations.

These episodes are rendered without technical detail, which may frustrate those hoping for insights into the science. Take the definition of renormalization — the subject of two papers that Dyson

published in *Physical Review* in 1949, work for which many feel he should have received a Nobel Prize. Schewe describes it as a process that involves redefining an electron's charge and mass by using the “countervailing blizzard of electromagnetic fields” to neutralize infinities that kept cropping up in theories, making the electron “normal again, fit to participate in atomic society”. However, Schewe's decision to pitch his breezy account at the non-scientist is, I think, justifiable, as the story has already been covered by several technical histories of physics, such as Frank Close's *The Infinity Puzzle* (Oxford Univ. Press, 2011) and Silvan Schweber's *QED and the Men Who Made It* (Princeton Univ. Press, 1994).

One of the pleasures of Schewe's book is waiting for him to uncork the next flamboyant metaphor. Some are more successful than others. “Like Odysseus, Dyson had been entangled in a great war and afterward

Maverick Genius: The Pioneering Odyssey of Freeman Dyson

PHILLIP F. SCHEWE
Thomas Dunne Books:
2013. 352 pp. \$27.99

arrived safely in Ithaca,” Schewe writes of his subject's migration to Cornell. That made me smile. But comparing Dyson's mugging in Washington DC to Prince Andrei's wounding at the battle of Austerlitz in Tolstoy's *War and Peace* made me wince. And likening Dyson to the Beatles — and his papers on renormalization to *Please Please Me* — is embarrassing.

A few of Schewe's metaphors about other elements of the story also misfire. He compares the twentieth-century physics community to the early Christian church, and calls the initial Shelter Island conference — the first important post-war US gathering of theoretical physicists — an “ecumenical council”. The conference was no such thing. Not all key figures attended it, and those who did reviewed and prepared, rather than decreed. It would be more accurate, and equally accessible, to compare the event to a symphony orchestra's pre-performance rehearsal.

Schewe doesn't shy away from the personal, such as who slept with whom in Princeton's intense, tightly knit community. Yet avoiding technical details sometimes inhibits his focus. From today's perspective, was TRIGA safe or risky? Was Orion realistic or foolhardy? Surely the answers matter and reveal aspects of Dyson the man, but Schewe doesn't give us enough information to judge.

In *The Scientist as Rebel*, one of many articles he has written for the *New York Review of Books*, Dyson himself writes that “science is a human activity, and the best way to understand it is to understand the individual human beings who practice it”. But sometimes an understanding of science is needed to understand the human being who engages in it. This is especially true for someone like Dyson, who uses scientific arguments to defend offbeat positions.

For instance, is Dyson's high-profile, controversial position on climate change — in which he challenges the need to take drastic measures — realistic or reckless? We cannot judge without more discussion of the science of modelling. Discussion of Dyson's opposition to the Superconducting Super Collider and to the Hubble Space Telescope — in his view, oversized enterprises that monopolize resources at the expense of numerous smaller, more worthy projects — would also have brought a sharper, more critical focus on Dyson's contrarian personality, and made the end of Schewe's book read less like an extended, flattering magazine portrait of a prophet. Still, *Maverick Genius* provides a vivid and enjoyable sketch of one of the most prominent scientific rebels of our day. ■

Robert P. Crease is professor of philosophy at Stony Brook University, New York 11794, USA, and author of *World in the Balance*. e-mail: robert.crease@stonybrook.edu



CONTRASTO/EVEMINE

➔ **NATURE.COM**
For Freeman Dyson
on Enrico Fermi, see:
go.nature.com/wq6jau



CHUCK DOSWELL

Two tornadoes near Zurich, Kansas. The central region of neighbouring Oklahoma is the most tornado-prone site in the world.

METEOROLOGY

On the twister trail

Chuck Doswell enjoys a history of the passionate pioneers behind tornado science.

Today, most 'tornado chasers' are hobbyists who literally chase, observe and photograph tornadoes in the central plains of the United States, popularly known as Tornado Alley. Lee Sandlin's chronicle *Storm Kings* features the country's first tornado hunters — its pioneering meteorologists. Although they lacked the technology for the physical chase, they shared a passion for understanding tornadoes, a feeling I share.

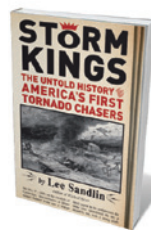
We now know that most tornadoes form from rotating storms called supercells, and result from a concentration of that rotation — in essence, what happens when a vortex forms in the water draining from a bathtub. The challenge for scientists today is to understand why so few storms produce tornadoes, and whether climate change will influence the frequency and intensity of these events.

From the early eighteenth to the early twentieth centuries, however — the period that dominates *Storm Kings* — the science of storms was in its infancy. Sandlin opens with US scientist and statesman Benjamin Franklin. Although best known for his investigation into the electrical nature of lightning,

Franklin was also interested in tornadoes. He witnessed one in Maryland in 1754, recognizing the rotation of its whirling winds.

Sandlin then jumps to the mid-nineteenth century and the vitriolic clash between James Espy, who developed a 'thermal inflow' theory of cyclones — in which warm air rises and is replaced by air flowing inward — and William Redfield, who believed that cyclones comprised rotating winds. This 'weather war' raged until Redfield's death in 1857. It is now clear that both were correct up to a point, a helpful reminder that in scientific controversy opponents may be right about some things and wrong about others.

I particularly enjoyed Sandlin's review of the pioneering efforts of John Park Finley



Storm Kings: The Untold History of America's First Tornado Chasers

LEE SANDLIN
Pantheon: 2013.
304 pp. \$26.95

(1854–1943), the first meteorologist to study tornadoes in any detail. Under the aegis of the US Army, Finley created the first tornado-spotter networks; wrote *Tornadoes* (1887), the first book on the subject; promoted the digging of underground tornado shelters; and issued the first experimental tornado forecasts and developed the first verification system to evaluate them. Finley's superiors eventually stopped his studies, which interfered with his military obligations. Subsequently, the word 'tornado' was banned in US public weather forecasts, as it was thought to cause public panic. The ban was lifted only in the early 1950s.

The quantitative risk of being hit by a violent tornado, even in central Oklahoma — the location with the highest frequency of them in the world — is so low that for any particular location, 1,000 years could pass without one. Yet someone, somewhere in the United States is hit every year. Finley understood the advisability of preparedness, and his view was validated in 2011 by the deaths of 553 people — the largest annual toll since 1925, when the Tri-State tornado devastated Missouri,

Illinois and Indiana, as Sandlin describes.

Storm Kings touches briefly on Robert Miller and Ernest Fawbush, US Air Force officers who issued the first post-Finley tornado forecast, for Tinker Air Force Base in central Oklahoma on 25 March 1948. There is also an introduction to the groundbreaking tornado research of Theodore Fujita, and an overview of the era of modern storm chasing.

The United States, with the world's highest tornado frequency, is a mecca for such hobbyists. But tornadoes have been observed on every continent except Antarctica. The vulnerability of populations around the world to tornadoes is complex, involving factors such as population density, construction standards and storm severity, as well as frequency. In Bangladesh, for instance, single tornadoes have had very high death tolls: one, which hit Daulatpur and Sattura on 26 April 1989, killed an estimated 1,300 people.

Sandlin's "storm kings" had an altogether simpler view of the risks, and were unaware of how tornadoes form in the outer rainbands of tropical cyclones or hurricanes. The potential for tornadoes to form as hurricanes make landfall is a real threat, but they tend to cause less damage than storm surges, which can wipe out miles of low-lying coastal areas.

We have also moved on in terms of our understanding of the psychological and behavioural aspects of preparedness: people have to believe that tornadoes might affect them personally. Preparation is not necessarily costly (an action plan and a stocked disaster kit are the basics), but attitude is key. Much remains to be learned about storms, not just about meteorology.

Storm Kings does contain technical errors and misconceptions. To take just one example, in discussing Espy's odd ideas about how water-vapour condensation powers thunderstorms, Sandlin mentions how temperature and pressure fall off with altitude. He refers to the calculation as "the saturated adiabatic lapse rate", but this term refers to the changes that occur under certain idealized assumptions, rather than real-world changes.

I have been a meteorologist interested in tornadoes for my entire career. Minor quibbles aside, I found *Storm Kings* a compelling history. ■

Chuck Doswell is Senior Research Scientist with the Cooperative Institute for Mesoscale Meteorological Studies affiliated with the University of Oklahoma. As part of his work as a consulting meteorologist, he has developed training materials for the National Weather Service and storm spotters. e-mail: cdoswell@earthlink.net

ARTS

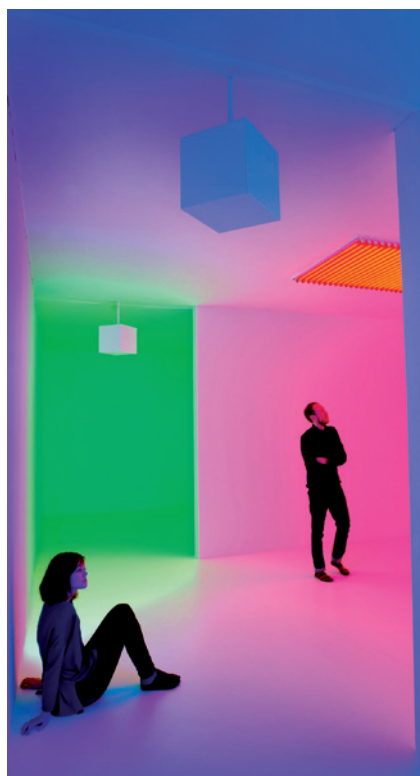
Feel the photons

Joanne Baker revels in a show that celebrates the sensory power of light.

I am in a pitch-black corridor, my senses straining. Feeling the wall as a guide, I take hesitant steps until I become aware of a dim glow. A shimmering wall of red bordered by a strip of pale turquoise beckons. Three radiant planes, set back at angles to one another and washed in shades of tomato and aquamarine, smack of abstract artist Mark Rothko's translucent layers. This is a painting in pure light.

James Turrell's engrossing *Wedgework V* (1974), is a highlight of *Light Show*, an exhibition exploring the qualities of light at London's Hayward Gallery. From mathematical shapes traced in fluorescent tubes to an unlit incandescent bulb whose mirror-reflected filament is — puzzlingly — most definitely on, works by 22 artists grapple with light's ability to dazzle and deceive.

Perfectly paced strobe lighting freezes the motions of 27 tabletop fountains in Danish-Icelandic artist Olafur Eliasson's *Model for a Timeless Garden* (2011). The arcing and burbling jets make a landscape of mercuric sculptures, organic forms that mesmerize as they slowly alter.



Chromosaturations by Carlos Cruz-Diez.

Shadows inspire Conrad Shawcross's sculpture *Slow Arc Inside a Cube IV* (2009). A shifting lat-

Light Show
HAYWARD GALLERY
Southbank Centre,
London.
Until 28 April 2013

tice of darkness cast on the gallery walls by a lamp moving within a wire cage is hypnotic, setting you off balance. Shawcross's work is influenced by chemist Dorothy Hodgkin, who described deducing protein structures using X-ray crystallography as akin to decoding the shape of a tree from the shadows of its leaves.

Plenty of the artists recreate natural illumination: a Las Vegas sunset captured in light boxes refashioned from shop signs; the spectrum of moonlight emitted by a bespoke halogen bulb; the phosphorescent flicker of cascades of light-emitting diodes (LEDs), which conjure up meteor showers or fireflies. Nancy Holt's row of circular perforations in a wall illuminated alternately from one side or the other recalls projections of the Sun's face or the transit of Venus across it.

Many installations are immersive. Visitors must don plastic shoe covers to step into a series of spaces bathed solely in blue, red or green. The monochromatic experience is disorienting, explains Venezuelan artist Carlos Cruz-Diez, because our retinas are used to perceiving a range of colours. Cone cells in the eye pick out each primary hue and other cells highlight opposing shades. Spend much time in a red room, and you start to appreciate even a tiny fleck of blue and green.

Theatrical lighting is used to great effect in Anthony McCall's *You and I, Horizontal* (2005). In a darkened room, puffs of mist bring the appearance of solidity to curved sheets of intense white light. Emanating from a point like an old film projector, the narrow sheets carve out a horizontal cone. Moving through this 'solid light' sculpture feels strangely subversive, like stepping through a wall.

Light Show does not expose much of the physical or optical qualities of light, as artists such as Eliasson have done elsewhere in works using prisms and optics. Revel instead in a playful and interactive look at the sensory side of light that illuminates as it entertains. ■

Joanne Baker is Senior Comment Editor at *Nature* in London.
e-mail: j.baker@nature.com

Correspondence

Scientists' rituals are ripe for investigation

I suspect that many aspects of the ritual behaviour being investigated by anthropologist Harvey Whitehouse and his colleagues (*Nature* **493**, 470–472; 2013) could well apply to academia, and to the scientific community in particular.

Science generally operates in what Whitehouse calls the 'doctrinal mode', but the 'imagistic mode' might also be relevant. Examples of the doctrinal mode include the ritual of the weekly lab meeting, the bonding induced by social outings, the ritualized nature of scientific conferences and the stereotypical behaviour of different lab members as they go about their research.

For the imagistic mode, what about the agony or ecstasy of having your paper rejected or accepted by *Nature*? Or the terrified first-year PhD student who has his or her results lambasted at a lab meeting or, worse, at a conference? It is not unknown for people to crumple under the onslaught of such a "traumatic ritual".

Studying the scientific community would be less dangerous than spending seven months in the middle of a civil war, and might encourage scientists to consider their own ritual behaviours.

Denis R. Alexander *The Faraday Institute for Science and Religion, Cambridge, UK.*
dra24@hermes.cam.ac.uk

Inertia is speeding fish-stock declines

A bluefin tuna weighing 222 kilograms sold for ¥155 million (US\$1.7 million) in Tokyo last month. This is an all-time record price for a single fish, driven skywards by increasing demand and dwindling supply.

Populations of bluefin tuna (*Thunnus* spp.) are a fraction

of what they were 50 years ago (see, for example, go.nature.com/5advqj), in large part because of the increasing value of the high-end sushi and sashimi market. Individuals are caught long before they reach full size (around 650 kg).

Conservation is being delayed as economic and political interests override biological concerns. The Convention on International Trade in Endangered Species of Wild Flora and Fauna helps to protect imperilled species, but has evaluated barely 3% of described species since it came into force in 1975. Similarly, the International Union for Conservation of Nature (IUCN) lists fewer than 100 fish species out of a total of 25,000 in any of the categories of its Red List of threatened species.

Even when complete information is available for heavily affected marine species (giant flatfish and billfish, for example), the IUCN's worldwide voting procedures impose delays on ranking in the list.

Kelly Swing *Tiputini Biodiversity Station, University San Francisco de Quito, Quito, Ecuador.*
kswing@usfq.edu.ec

Iran needs global support to fight HIV

HIV infection rates have halved over the past decade in much of the developing world. In the Middle East, meanwhile, they have increased by more than 35%. There are 96,000 people living with HIV/AIDS in Iran alone, with new infections increasing most rapidly among women. Efforts to control the virus are starting to pay off, but international support is needed to sustain these encouraging results.

Iranian scientists have developed a useful model for treating drug addiction and preventing transmission of HIV (C. S. Todd *et al.* *Curr. HIV/AIDS Rep.* **4**, 151–157; 2007)

that has brought the spread of HIV among injecting drug users under control. But Iran will need to adapt its response as the epidemic shifts to new populations.

HIV researchers, clinicians and policy-makers from the United States, Europe and the Middle East met in Tehran last year to identify priority targets. These include preventing mother-to-child transmission, expanding treatment programmes and reducing infection risk in key populations (see iranhivandaidscongress.org).

International collaborations will help Iran to meet these public-health challenges.

Kayvon Modjarrad *National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA.*

kayvon.modjarrad@nih.gov

Minoo Mohraz *Tehran University of Medical Sciences, Tehran, Iran.*

Navid Madani *Dana-Farber Cancer Institute, Boston, Massachusetts, USA.*

Educate physicians about investor types

As a venture capitalist at a US investment fund that supports young pharmaceutical companies, I believe that the misuse of clinical experts by a few hedge-fund investors should not prevent interaction with physicians (*Nature* **493**, 271–272, 280–281; 2013). Increasing regulation would risk stigmatizing physician participation in such expert networks, stifling information flow for investors and ultimately impeding drug discovery.

Academic and hospital conflict-of-interest policies that target or ban engagement with investment firms would need to differentiate between investor types. Hedge-fund investors operate in public markets and aim for a quick profit by betting on variables such as clinical-trial

data, sales announcements or regulatory approval. Venture capitalists, by contrast, operate in private markets and aim for company growth over several years before selling to a large multinational or to investors through a public stock-market offering.

Unlike hedge-fund investors, venture capitalists holding public securities as insiders cannot profit by selling stock before the release of data. Furthermore, their interests are not threatened by confidential information leaks because of the longer-term nature of their investments.

It would be preferable for investors to identify conflicts of interest with respect to specific drugs before interviewing physicians, and to educate these experts formally on the potential pitfalls of engaging with investors.

Justin Chakma Thomas, *McNerney and Partners, La Jolla, California, USA.*
jchakma@tm-partners.com

Fungus did not cause potato famine

The agent responsible for the blight that caused the nineteenth-century Irish potato famine, *Phytophthora infestans*, should not be "grouped with fungi" (*Nature* **493**, 154–156; 2013).

It was Anton de Bary, the father of mycology, who coined the genus name *Phytophthora* ('plant-destroyer') and classed the pathogen as a fungus. But modern molecular sequencing indicates that his interpretation was incorrect (M. D. M. Jones *et al.* *Nature* **474**, 200–203; 2011).

The organism is actually an oomycete, a pseudo-fungus that evolved from killer ancestors in the ancient oceans and not from wood-degrading fungi.

U. Kutschera *Institute of Biology, University of Kassel, Kassel, Germany.*
kut@uni-kassel.de

Of humans, dogs and tiny tools

Genomic data hint at the possibility of human migration from India to Australia 4,230 years ago. However, the inference that these humans took along their dogs and tools is difficult to reconcile with previous reports.

PETER BROWN

Reporting in *Proceedings of the National Academy of Sciences*, Pugach and colleagues¹ provide genetic evidence of a possible mid-Holocene (4,230 years ago) link between human populations in India and Australia. Their data confirm the results of some genetic studies², but contradict others³. Intriguingly, the authors also link this evidence to the arrival of the dingo and the appearance of microlithic stone tools, which appeared in India as early as 34,000 years ago⁴, but much more recently in Australia.

The fascination with human migration to Australia began in the eighteenth century, when European explorers reached its coastline. They were surprised to find indigenous human inhabitants and dogs on a continent that was otherwise filled with alien flora and fauna. Australia was separated from both the Asian mainland and the Indonesian archipelago by sea, so where had Aboriginal Australians and dogs come from, and when?

Unfortunately, for succeeding generations of anthropologists and archaeologists, determining the origins of these early navigators remained a vexatious issue. Until recently, all that could be said with confidence was that by 45,000 years ago modern humans had occupied large tracts of Australia⁵, that water craft were required to reach the Sahul Shelf from Asia, and that for most of its human history Australia was a relatively isolated place^{3,6}. Issues relating to the dispersion or the independent invention of aspects of material culture (rock art, ground-edge axes and microlithic tools), as well as to the spread of dingoes, which could shed light on human origins and cultural development, remained unresolved.

Earlier studies^{2,3} have highlighted the ancient roots of the first Australians, but have disagreed over the evidence of Indian contact. In one case, Y-chromosome analysis² indicated tight links to Indian and Sri Lankan lineages in the mid-Holocene, but another³ study of Y chromosomes and DNA from cellular organelles called mitochondria found no evidence of a specific



Figure 1 | Sahul Shelf cousins. Three-dimensional reconstructions from microtomography scans of the crania of a female dingo and New Guinea singing dog. Although different in size, the two species are genetically closely related and exhibit similar behavioural and anatomical features. In contrast to Pugach and colleagues' suggestion¹, dingo ancestors probably arrived in Australia from Papua New Guinea. Scale bar, 50 millimetres. (Scans by Richard Flavel, Univ. New England.)

relationship with South Asian lineages.

Pugach *et al.* approached these issues by analysing single nucleotide polymorphisms (SNPs) across autosomal (non-sex) chromosomes of 344 individuals, including Aboriginal Australians, New Guineans, island Southeast Asians and Indians, as well as those of European and Chinese origin. SNPs describe DNA-sequence variations that are present in at least 1% of members of a species⁷.

In agreement with previous studies^{2,3}, the authors report a close genetic relationship between Aboriginal Australians and both the highlanders of New Guinea (who are known to have an ancient common origin) and the Mamanwa — a Negrito group from the Philippines. The authors estimate the divergence time between these three groups to be 36,000 years ago, which is too recent to tally with the archaeological evidence of dispersal onto the Sahul Shelf about 45,000 years ago. The authors acknowledge that the method they used to estimate divergence times may have produced an underestimate.

What makes this research contentious is that it indicates mid-Holocene gene flow from

India to Australia. Pugach *et al.* rule out the possibility that this signal in their Aboriginal sample is a result of the recent European genetic admixture. Instead, using simulations, they pinpoint the Indian admixture time to 141 generations ago, which — with a generation time of 30 years — dates the gene flow to 4,230 years ago.

The genetic evidence of a possible prehistoric link between Australia and India is intriguing, particularly given the previously accepted genetic and archaeological evidence for isolation^{3,8}. Until now, the most substantial evidence of pre-European contact was with Macassan trepang (sea cucumber) fishermen, who sailed to northern Australia from around 1640. Indeed, the significant cultural impact of these long-term contacts on Aboriginal Australians is evident in their language, art and material culture. Surely, a substantial genetic signal from presumably Neolithic navigators should be similarly reflected in evidence of cultural transmission. So I imagine that many archaeologists and biologists who work in the Australasian region would dispute Pugach and colleagues' suggestion

that this possible India-to-Australia gene flow is connected with the spread of the dingo and microlith use.

The dingo has only ever been found on the Australian mainland, presumably arriving some time after post-glacial flooding of the Bass Strait around 12,000 years ago. The earliest reliably dated evidence of the dingo is about 3,450 years old from Madura Cave — roughly 800 years after the suggested Indian contact. Given that the introduced European fox crossed the continent in only 80 years, you might expect closer correspondence between the inferred dates of introduction and the palaeontological and archaeological evidence.

As for the origin of these animals, when mitochondrial DNA from 211 dingoes was compared with that of 676 dogs from all continents⁹, dingoes were shown to originate from domesticated East Asian — not Indian — dogs, and may have been part of the Austronesian expansion into Southeast Asia around 4,500 years ago. Specifically, dingoes were found to be closely related to the much smaller New Guinea singing dog (Fig. 1), and both species were shown to have descended from an early East Asian domesticated dog. The ancestors of living dingoes, therefore, may have been introduced across the Torres Strait, which separates Papua New Guinea from Australia, at some point after it flooded 8,000 years ago.

The arrival of microlithic tools in Australia as part of a package with the dingo and contact with Indian populations may be equally unlikely. Although it was once thought that the large-scale production of microliths dated to the period after 4,000 years ago, corresponding with the earliest evidence of the dingo, there is strong evidence that microliths were in use many millennia before the dingo appeared⁸ — for instance, 6,000 years ago at Capertee 3 rockshelter and possibly as early as 15,500 years ago at Walkunder Arch.

Unfortunately, there is almost no archaeological evidence that modern humans were in Asia before their first appearance in Australia. Ultimately, the resolution of questions that surround the origins of Australia's first human colonists — and any subsequent contact with maritime peoples to the north — will come from a synthesis of genetic, archaeological and palaeoanthropological data, which have been used to estimate the time taken for the expansion of Austronesian speakers from Asia into the Pacific. ■

Peter Brown is in the Faculty of Arts and Sciences, University of New England, New South Wales 2351, Australia.
e-mail: pbrown3@une.edu.au

- 8726–8730 (2007).
4. Clarkson, C. *et al. Antiquity* **83**, 326–348 (2009).
5. O'Connell, J. F. & Allen, J. *Evol. Anthropol.* **6**, 132–146 (1998).
6. Rasmussen, M. *et al. Science* **334**, 94–98 (2011).
7. Barreiro, L. B., Laval, G., Quach, H., Patin, E. &

- Quintana-Murci, L. *Nature Genet.* **40**, 340–345 (2008).
8. Hiscock, P. *Archaeology of Ancient Australia* (Routledge, 2007).
9. Savolainen, P., Leitner, T., Wilton, A. N., Matisoo-Smith, E. & Lundeberg, J. *Proc. Natl Acad. Sci. USA* **101**, 12387–12390 (2004).

AGEING

Stem cells on a stress-busting diet

Knowing how an organism's tissues handle stress throughout life is key to understanding ageing and disease. Stem cells of the blood system seem to tackle metabolic stress by means of a process called autophagy. [SEE ARTICLE P.323](#)

TERESA V. BOWMAN & LEONARD I. ZON

Stem cells in adult tissues function to replace lost, damaged or diseased cells throughout an organism's life, thereby helping to maintain tissue health. But what protects the rare, long-lived stem cells from a lifetime of exposure to cellular and environmental stressors such as inflammation, radiation and metabolic alterations? On page 323 of this issue, Warr *et al.*¹ propose that autophagy, a process of cellular self-cannibalization, is one mechanism that haematopoietic stem cells of the blood system use to protect themselves during times of metabolic stress, when nutrients are limited*.

When cells are starved of nutrients, the

*This article and the paper under discussion¹ were published online on 6 February 2013.

tissue's stem cells must choose whether to live or die. The options are committing suicide by apoptotic cell death, or self-preservation through autophagy, whereby cells recycle damaged or dispensable proteins and organelles into basic components to support cellular growth.

Autophagy is thought to be a major factor in ageing². Loss of autophagy in tissues such as the brain, liver and heart leads to an increase in age-related disorders, including neurodegeneration, metabolic syndromes and cardiac dysfunction. Conversely, factors that stimulate autophagy abrogate these problems and have been linked to greater longevity. It has therefore been hypothesized that reduced autophagy contributes to the diminishing stem-cell function that occurs with age. No study, however, has investigated the direct role

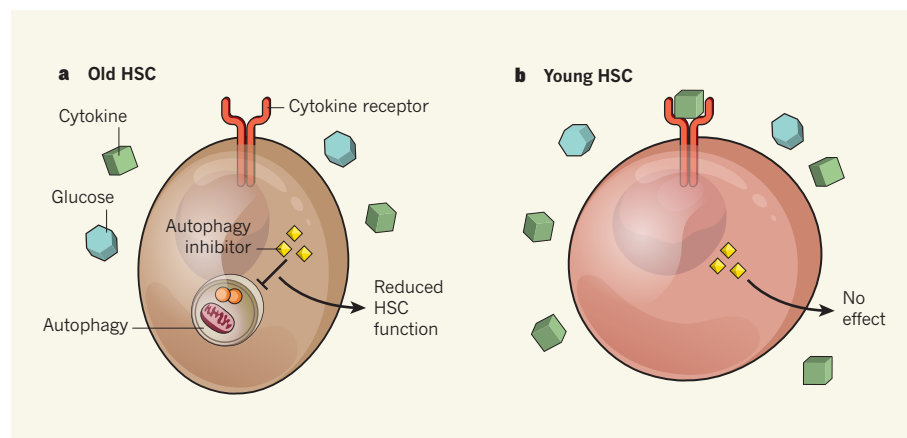


Figure 1 | Stem cells, nutrient uptake and autophagy. **a**, Warr *et al.*¹ report that old haematopoietic stem cells (HSCs) cannot take up glucose normally and compensate for the resulting metabolic stress by using the process of autophagy to break down their own components as a source of nutrients. Unsurprisingly, therefore, blocking autophagy in these cells impairs their function. The surrounding environment of old HSCs contains low levels of cytokines, which are known to mediate glucose uptake by binding to cytokine receptors on the surface of the cells. **b**, Young HSCs, by contrast, can take up glucose normally, perhaps because cytokine levels in their environment are high. These cells can thus support their own metabolic needs, and so blocking autophagy under non-stress conditions does not notably affect their function.

1. Pugach, I., Delfin, F., Gunnarsdottir, E., Kayser, M. & Stoneking, M. *Proc. Natl Acad. Sci. USA* **110**, 1803–1808 (2013).
2. Redd, A. J. *et al. Curr. Biol.* **12**, 673–677 (2002).
3. Hudjashov, G. *et al. Proc. Natl Acad. Sci. USA* **104**,



50 Years Ago

The mathematician of to-day is a much more powerful figure in research than his predecessor of fifty years ago. Indeed, it is fair to say that the mathematician is gradually taking over much of applied physics. This need not worry the physicist, who is moving into chemistry with growing momentum. Nor the chemist, who, with the physicist, is now a key figure in biological research: and what of the biologist? Does he need to worry? The answer is surely no, for he will become more and more a leader in the sociological fields, ergonomics being a case in point.

From *Nature* 23 February 1963

100 Years Ago

Nutrition Physiology. By Prof. P. G. Stiles — It is not possible to regard the book as a mere addition to the already numerous primers of physiology; it is something beyond this, although it makes no pretensions to being anything profound. It can be read with profit by the junior student, and still more by the senior student, and even the professed physiologist. Old truths are often put in new ways, and so fresh light is shed upon familiar problems ... The book contains the inevitable chapter on alcohol; this is written in a moderate strain, and may, perhaps be viewed with disfavour by the extreme teetotaler because it is not intemperate. As one reads it, one almost feels that its author was writing it because he had to, but was protesting all the time inwardly against the American law which excludes all physiological books from scholastic institutions which do not obey the tyrannical behests of the party in power.

From *Nature* 20 February 1913

of autophagy in adult stem-cell function.

Warr *et al.* explore this question in both young mouse haematopoietic stem cells (HSCs) and in more-differentiated HSC progeny, including progenitor cells of the immune cells granulocytes and macrophages. The authors find that little or no autophagy occurs in freshly isolated young HSCs, but that this process can be rapidly induced when the cells are exposed to metabolic stress both *in vitro* and *in vivo*. Moreover, when autophagy is inhibited during such metabolic stress, young HSCs rapidly die through apoptosis, indicating that autophagy is crucial for their survival. By contrast, granulocyte-macrophage progenitor cells show higher baseline levels of autophagy, but no shift under starvation conditions.

Autophagy can be stimulated in several ways², including through inhibition of the signalling molecule mTOR and activation of stress-induced transcription factors such as FoxO3 and p53. Warr and colleagues find that the primary driver in HSCs is FoxO3, with little contribution from mTOR or p53.

FoxO3 is a member of the FoxO family of transcription factors, which are involved in diverse processes, including stress resistance, apoptosis and metabolism. Loss of a single type of FoxO protein has been found to have little effect on HSCs, but simultaneous loss of three of them (FoxO1, FoxO3 and FoxO4) leads to dramatic changes in HSC proliferation and in levels of reactive oxygen species³. Warr *et al.* propose that a similar redundancy may explain the incomplete loss of autophagy seen when only FoxO3 is missing. It is also possible that sirtuin proteins act upstream of FoxO3, as they are known to promote autophagy and increase longevity — in part by activating FoxO transcription factors — and to affect the proliferation of aged HSCs and their response to metabolic stress^{2,4,5}.

The implication of autophagy in ageing necessitates a better understanding of whether old stem cells use autophagy for similar purposes to their younger counterparts. Such knowledge should provide insight into the mechanisms underlying ageing and disease, and lead to improved strategies for enhancing health and longevity. Warr and co-authors show that, unlike young HSCs, freshly isolated old HSCs show some autophagy. Also, old cells could mount a similar response to their younger counterparts under starvation conditions. These findings contradict the currently held belief that autophagy diminishes with age.

Nonetheless, abrogated autophagy had different effects on the function of old and young HSCs. The authors report that blocking autophagy diminished the colony-forming potential of old HSCs to a much greater extent than in young cells. Colony formation is an *in vitro* metric used to determine the capacity of blood cells to form differentiated cell

types. So although old and young cells show the same levels of autophagy, old cells seem to be more dependent on autophagy for their functioning. Confirmation of this would be to show that loss of autophagy more severely affects the capacity of old than young HSCs to repopulate the blood system of an individual whose immune system has been destroyed by radiation — the gold standard for assessing the function of HSCs.

The differences in autophagy between old and young HSCs did not seem to be due to changes in the activities of FoxO3, p53 or mTOR, and the transcriptional program triggered by FoxO3 was still active in old cells. Surprisingly, however, the old cells showed defects in the uptake of glucose, and so were in a state of metabolic stress even under normal conditions (Fig. 1). Feeding old cells a nutrient source that bypassed the need for glucose uptake restored the cells' colony-forming ability even when autophagy was blocked.

A study in mice recently indicated⁶ that deregulation of cytokine proteins might be one explanation for the altered nutrient uptake in aged HSCs, as cytokines help to promote glucose uptake from the cellular environment. If such metabolic stress also occurs in humans, correcting it could help to improve stem-cell function in elderly people. Alternatively, experiencing prolonged autophagy may affect metabolism in old HSCs, as autophagy mediates the clearance of damaged mitochondria (organelles that act as cellular powerhouses), potentially resulting in lower energy output and slowing with age of essential cellular processes.

An intriguing question is whether the age-associated increase in autophagy is specific to HSCs or also occurs in stem cells of other tissues. So far, most of the other tissues studied for traits associated with age-related autophagy, such as the brain and heart, are those with a slow turnover of cells². It could be that stem cells in highly proliferative tissues such as the blood have a different dependence on autophagy. For a definitive answer to this, researchers should examine stem cells in other tissues with a high cell turnover, such as the skin and gut. ■

Teresa V. Bowman and Leonard I. Zon
are in the Stem Cell Program, Boston
Children's Hospital, Boston, Massachusetts
02115, USA.

e-mails: tvbowman@enders.tch.harvard.edu;
zon@enders.tch.harvard.edu

1. Warr, M. R. *et al.* *Nature* **494**, 323–327 (2013).
2. Rubinsztein, D. C., Mariño, G. & Kroemer, G. *Cell* **146**, 682–695 (2011).
3. Tothova, Z. *et al.* *Cell* **128**, 325–339 (2007).
4. Kroemer, G., Mariño, G. & Levine, B. *Mol. Cell* **40**, 280–293 (2010).
5. Ou, X. *et al.* *Blood* **117**, 440–450 (2011).
6. Ergen, A. V., Boles, N. C. & Goodell, M. A. *Blood* **119**, 2500–2509 (2012).

CLIMATE SCIENCE

Global warming and tropical carbon

An innovative use of measurements of atmospheric carbon dioxide constrains the possible range of carbon-cycle responses to climate change during the twenty-first century, lowering expectations of tropical-forest dieback. [SEE LETTER P.341](#)

JAMES T. RANDERSON

Tropical forests and savannahs are responsible for most of the annual exchange of carbon dioxide between the atmosphere and the land surface. Long growing seasons and high rainfall create extremely productive tropical ecosystems¹ that store a considerable amount of the world's above-ground biomass². The fate of this carbon pool could influence future atmospheric carbon levels and thus our ability to stabilize greenhouse gases at levels that do not dangerously interfere with the climate system³. On page 341 of this issue, Cox *et al.*⁴ report findings that will enable the development of more realistic future scenarios of climate-carbon interactions in tropical ecosystems*.

Over the past decade, tremendous progress has been made towards integrating the carbon cycle into global climate models known as Earth-system models. Simulations using such models have shown that carbon stocks in tropical forests will become increasingly vulnerable to climate change during the twenty-first century, given future 'business as usual' scenarios of fossil-fuel emissions. In one model, for example, climate warming and elevated levels of CO₂ reduced precipitation across the Amazon, which led to a die-off of forests and thus further warming⁵. In several other models, increases in drought stress in forests across the tropics decreased plant growth (net primary production), and so accelerated carbon losses from these ecosystems⁶.

These interactions, in which climate change induces a carbon-cycle change that contributes to additional warming, are called positive climate-carbon feedbacks. The strength of these feedbacks varied by more than a factor of seven for models analysed⁶ as part of the Intergovernmental Panel on Climate Change's Fourth Assessment Report. In most models, the response of tropical ecosystems dominates the strength of the global-scale climate-carbon feedbacks, although losses from high-latitude carbon pools are also important in models that incorporate the best current representation of permafrost-soil physics⁷.

Cox *et al.* have developed an approach for

reducing the considerable uncertainty that is associated with the size of tropical climate-carbon feedbacks. They used an 'emergent model constraint'⁸ to refine the probable range of sensitivities of tropical-land carbon fluxes to climate change during the twenty-first century. This approach has previously been used to constrain projections of climate change, and involves two steps. First, a statistically significant relationship must be found between contemporary and future model predictions for an ensemble of Earth-system models from different research centres. Second, observations must be used to limit the possible range of contemporary behaviours described by those models.

The authors show that this approach can be applied to the carbon cycle, and demonstrate that models that show higher levels of year-to-year (interannual) variability of atmospheric CO₂ levels in response to variations in climate

modes, such as the El Niño–Southern Oscillation, tend to predict larger losses of carbon from the tropics in response to twenty-first-century climate change. Cox *et al.* combined the obtained linear relationship with high-quality measurements of atmospheric CO₂ and temperatures from global observing networks, and narrowed the range of probable future outcomes. Their analysis led to a striking conclusion: many Earth-system models probably overestimate the magnitude of the tropical carbon losses that would occur in response to climate warming. As a result, the authors argue, the likelihood of a tropical dieback event is considerably lower than might be inferred from earlier work.

Does this mean that we no longer need to worry that climate change will bring about the loss of tropical-forest ecosystems? Not necessarily. Obtaining the correct magnitude of year-to-year atmospheric CO₂ responses to climate modes is necessary, but not sufficient, for accurate, longer-term carbon-cycle predictions over decades to centuries. For example, temperature changes and drought events that drive much of the interannual variability in contemporary atmospheric CO₂ concentrations have a fairly short duration because of the physical processes that regulate the El Niño–Southern Oscillation. These climate anomalies therefore have considerable immediate effects on 'fast' tropical-ecosystem processes, such as gross primary production, ecosystem respiration and fires. But they probably have



Figure 1 | Fire fighting in an Indonesian tropical forest. Cox *et al.*⁴ report that the likelihood of tropical forests dying off in response to global warming is much lower than previous studies have suggested. However, the effects of events such as tropical-forest fires must be integrated into climate models to assess more accurately the risk of dieback.

*This article and the paper under discussion⁴ were published online on 6 February 2013.

only moderate effects on tree recruitment, competition and mortality, which are likely to shape ecosystem responses to climate change on longer timescales. New observations of forest dynamics are needed for further advances, along with improved representation in Earth-system models of climate feedbacks that are associated with human activity, such as fires⁹ (Fig. 1) and land-use change¹⁰.

Cox and colleagues' approach is exciting because it opens up fresh avenues for evaluating Earth-system models, and provides opportunities to reduce the uncertainties that are associated with future scenarios of global environmental change. Their study also provides clear motivation for scientists who study

the carbon cycle to develop more sophisticated approaches for parsing contributions to variability in contemporary atmospheric CO₂ that arise from changes in photosynthesis, fires, tree mortality, respiration from plants and microbes, and rates of ocean upwelling. The sensitivity of each of these processes to long-term climate change deserves further inquiry, both through mechanistic studies and through the discovery of additional emergent constraints, following Cox and co-workers' lead. ■

James T. Randerson is in the Department of Earth System Science, University of California, Irvine, California 92697, USA.
e-mail: jranders@uci.edu

1. Nemani, R. R. *et al. Science* **300**, 1560–1563 (2003).
2. Saatchi, S. S. *et al. Proc. Natl Acad. Sci. USA* **108**, 9899–9904 (2011).
3. Solomon, S., Plattner, G. K., Knutti, R. & Friedlingstein, P. *Proc. Natl Acad. Sci. USA* **106**, 1704–1709 (2009).
4. Cox, P. M. *et al. Nature* **494**, 341–344 (2013).
5. Betts, R. A. *et al. Theor. Appl. Climatol.* **78**, 157–175 (2004).
6. Friedlingstein, P. *et al. J. Clim.* **19**, 3337–3353 (2006).
7. Koven, C. D. *et al. Proc. Natl Acad. Sci. USA* **108**, 14769–14774 (2011).
8. Hall, A. & Qu, X. *Geophys. Res. Lett.* **33**, L03502 (2006).
9. van der Werf, G. R. *et al. Proc. Natl Acad. Sci. USA* **105**, 20350–20355 (2008).
10. Davidson, E. A. *et al. Nature* **481**, 321–328 (2012).

MICROBIAL OCEANOGRAPHY

Killers of the winners

Viruses that infect the SAR11 group of oceanic bacteria have finally been found and sequenced. Because SAR11 is ubiquitous, these viruses may be the most abundant in the oceans — and perhaps in the entire biosphere. SEE LETTER P.357

DAVID L. KIRCHMAN

More than 20 years ago, the discovery of billions of viruses in the oceans was big news, worthy of articles in *Nature*¹ and on the front page of the *Washington Post*. A year later another *Nature* report was published², this time about the most abundant bacterial group in the oceans, cryptically called SAR11. The two stories now come together in a paper on page 357 of this issue. Zhao *et al.*³ describe DNA viruses that they call 'pelagiphages' and which infect laboratory-grown representatives of SAR11 bacteria

(Fig. 1). The authors use genomic-sequence data to argue that pelagiphages are among the most abundant viruses in the oceans and perhaps the entire biosphere. The report ends long-running speculation about SAR11, but prompts new questions about marine viruses and the control of microbes in the oceans⁴.

The general importance of bacteria in marine ecosystems was recognized long before the discovery of SAR11. Bacteria were known to be the most abundant organisms in the oceans — numbering nearly 10⁹ per

*This article and the paper under discussion³ were published online on 13 February 2013.

litre of surface sea water — and to be mediators of many biogeochemical processes in the cycling of carbon and other elements. However, until about 1990, microbiologists had identified only a few marine bacterial species because most could not be isolated and cultivated in the lab using the methods available at the time.

The development of cultivation-independent methods based on sequencing the 16S ribosomal RNA gene, which contains taxon-specific signatures, markedly enhanced our ability to distinguish between types of bacteria. These methods enabled the discovery of SAR11 bacteria in the Sargasso Sea² (hence the SAR moniker) and eventually their identification in every other marine system examined. There is even a relative of SAR11 in freshwater environments⁴. The bacteria's preferred habitat is the surface layer of the pelagic zone (water of oceans and lakes away from the shore and bottom), where the bacteria are so ubiquitous that when a SAR11 strain was finally isolated and cultivated in the lab it was dubbed *Pelagibacter ubique*⁵. This name was the inspiration for pelagiphage — the phages, or viruses of bacteria, found by Zhao and colleagues.

With the discovery of SAR11's abundance came the obvious question of why. What makes this bacterial clade so successful? Part of the answer is that the group is diverse, with different SAR11 types dominating different oceanic waters. But other bacterial groups are diverse, so why is SAR11 the clear winner in the oceans? One popular explanation, based on the small cell size and slow growth of *P. ubique*, is that SAR11 differs from other bacteria by being a defence specialist that fends off viruses and thereby achieves high abundance despite slow growth. This hypothesis was supported by the apparent lack of a SAR11-attacking virus, even though being perfectly impervious to viruses would make SAR11 unique in the biosphere.

In fact, the negative evidence was never very convincing, given the difficulties of identifying viruses in nature. The usual approach in virus characterization is to start by propagating the viruses in hosts grown on solid media,

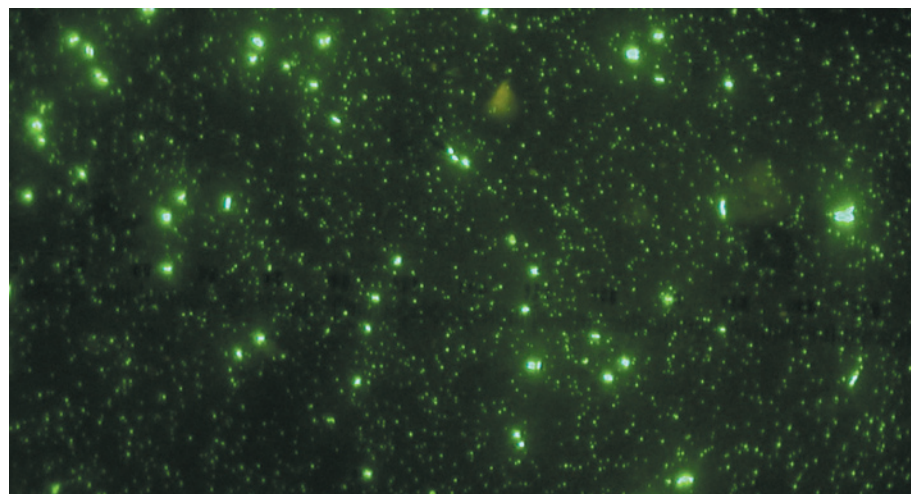


Figure 1 | Ocean inhabitants. A photomicrograph of a seawater sample taken off the coast of California. The larger dots are bacteria (about 0.5 micrometres in diameter) and the smaller ones are viruses; both are stained with the DNA-specific stain SYBR Green. The bacteria predominantly belong to the SAR11 group. Zhao and colleagues' results³ suggest that many of the viruses are pelagiphages that infect SAR11 bacteria.

JED FUHRMAN

but this form of cultivation is not possible for most abundant bacteria in nature. So, no host, no virus. Even when a method of culturing *P. ubique* was developed, finding viruses that might infect it was not straightforward because this SAR11 representative grows only in liquid media. In addition, the sequencing approaches that are used to identify bacteria do not work for viruses because they lack a name-tag gene such as the 16S rRNA gene.

To solve this identification problem, Zhao *et al.* sequenced four pelagiphages and compared their genomes with those of known and unidentified viruses. The sequence data, as well as the morphology of the viruses, indicated that pelagiphages belong to the same families as the viruses that attack cyanobacteria. The authors then estimated the relative abundance of SAR11 viruses by combining the pelagiphage genomic data with viral metagenomic data from several oceans, including the Pacific Ocean. (Metagenomic data are taken directly from all targeted organisms or viruses in a sample without prior cultivation.) The analyses suggest that SAR11 viruses are more numerous than viruses known to attack marine bacteria from three other abundant groups: *Roseobacter*, *Prochlorococcus* and *Synechococcus*. Zhao *et al.* successfully defend their pelagiphage numbers by considering methodological problems,

including 'greedy' recruitment. One unsolved issue is that the authors' calculations depend on data from viruses that infect just a few bacterial types — a puny number compared with the huge diversity of marine microbial communities. Still, they have identified a good chunk of the even larger genomic diversity carried by marine viruses.

Zhao *et al.* conclude by considering what implications their data have for the popular 'kill-the-winner' hypothesis — which asserts that as a bacterial type increases in abundance, it should attract more viruses, preventing it from becoming dominant⁶. The discovery of pelagiphages suggests this is not the case for SAR11, but the existence of SAR11 viruses should not be a surprise. SAR11 could still be a defence specialist, albeit an imperfect one. Indeed, Zhao *et al.* discuss signs in the *P. ubique* genome of defences against viruses. The 'defence specialist' label may fit if SAR11 grows as slowly in the oceans as *P. ubique* does in the lab. However, although some studies suggest slower than average growth rates for SAR11 (ref. 7), others have found faster rates⁸. The latter seems consistent with the superior competitive traits of SAR11 suggested by *P. ubique* strains and their genomes.

Determining whether SAR11 is a defence specialist or a superior competitor is important. The answer would help to explain

SAR11's population dynamics and be a notable start in determining the bacteria's significance in the carbon cycle and other biogeochemical processes. A slow-growing defence specialist may be on the sidelines of the biogeochemical action carried out by faster-growing competitors. Zhao *et al.* touch on these issues, but more work is needed. Nevertheless, by uniting two stories that started more than 20 years ago, this report is a notable chapter in understanding the most abundant viruses and bacteria in the oceans. ■

David L. Kirchman is in the School of Marine Science and Policy, University of Delaware, Lewes, Delaware 19958, USA.
e-mail: kirchman@udel.edu

- Bergh, Ø., Børsheim, K. Y., Bratbak, G. & Heldal, M. *Nature* **340**, 467–468 (1989).
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. *Nature* **345**, 60–63 (1990).
- Zhao, Y. *et al.* *Nature* **494**, 357–360 (2013).
- Bahr, M., Hobbie, J. E. & Sogin, M. L. *Aquat. Microb. Ecol.* **11**, 271–277 (1996).
- Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. *Nature* **418**, 630–633 (2002).
- Thingstad, T. F. & Lignell, R. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).
- Campbell, B. J., Yu, L., Heidelberg, J. F. & Kirchman, D. L. *Proc. Natl Acad. Sci. USA* **108**, 12776–12781 (2011).
- Malmstrom, R. R., Cottrell, M. T., Elifant, H. & Kirchman, D. L. *Appl. Environ. Microbiol.* **71**, 2979–2986 (2005).

INFECTION BIOLOGY

Cheats never prosper

Fast-growing 'defector mutants' can threaten the success of a bacterial infection. But one bacterial species prevails over these cheats by forming a subpopulation that has shut down expression of virulence genes. SEE LETTER P.353

DAVID T. MULDER & BRIAN K. COOMBS

Bacterial pathogens promote their survival by deploying virulence factors that modify the host environment. But these virulence proteins are energetically costly to produce, and so avirulent defector mutants that grow more quickly than the wild-type population can compromise the success of an infection. On page 353, Diard *et al.*¹ report a remarkable strategy used by the gastrointestinal pathogen *Salmonella enterica* serovar Typhimurium (hereafter termed *Salmonella typhimurium*) to evade this threat, while at the same time safeguarding its genotype — it produces a subpopulation of equally fast-growing bacteria that are rendered phenotypically avirulent through regulation of gene expression rather than mutation.

When studying host–pathogen interactions, a common misconception is that the bacterial population that initiates the infection remains

as a genetically identical collective of cells that retains its initial characteristics, or phenotype. In fact, most *in vivo* data on virulence-factor function and bacterial fitness are inferred from population-level studies, which do not consider the fate of individual bacteria. However, studies of antibiotic-resistant persister cells and the social evolution of cooperative traits in bacteria² have reinforced the idea that regulation of gene expression is a significant source of phenotypic variation within a population. For example, in a process called bistability, a single bacterial population can bifurcate into two subpopulations that are genetically identical, but phenotypically different. Bistability is epigenetic in the sense that the subpopulations arise without hereditary changes to the DNA sequence, but instead through changes to gene expression. The concept is not new^{3,4}, but the evolutionary significance of bistability for pathogenic bacteria during infection of a host has remained unclear, until now.

A well-known example of bistability is the expression of virulence factors by *S. typhimurium*. These bacteria deploy an extensive suite of virulence proteins that generates an inflammatory response in the infected host and allows the bacteria to invade the epithelial cells that line the gut (Fig. 1). The inflammation also kills off some of the non-pathogenic commensal bacteria that normally reside in the gut, creating an expanded niche in which *S. typhimurium* can grow⁵. Virulence factors are known to be expressed in a bistable way by the *S. typhimurium* population⁶, which generates subpopulations of slow-growing virulent cells that perform 'public good' functions that benefit the whole population, and fast-growing cells that do not express virulence factors but are beneficiaries of this public good⁷. This provides a fitness advantage to avirulent, non-cooperating bacteria, but it also raises the question of whether the infecting population is vulnerable to invasion by avirulent mutant defectors. Although the susceptibility that results from this social structure has already been firmly established^{8,9}, Diard and colleagues have provided greater clarity by analysing the genetics of the within-host evolution that takes place during infection.

To better understand the altruistic cooperative behaviour of the virulent subpopulation, the authors modelled the effect of various infection ratios of cooperating bacteria and mutant cheats, and tested these scenarios *in vivo* in

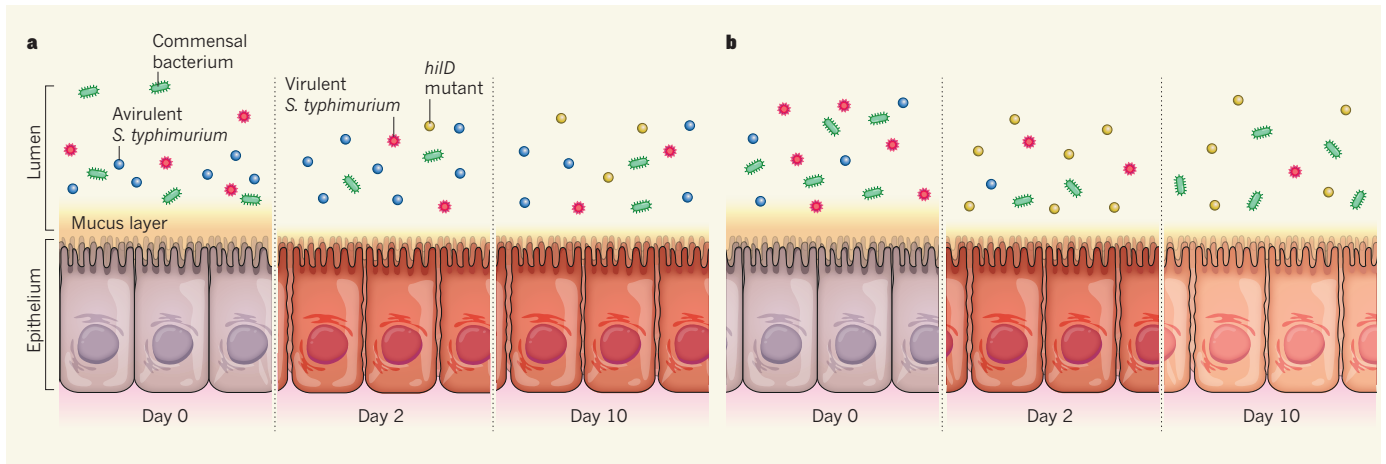


Figure 1 | Bistability prevents the rise of mutant defectors. Diard *et al.*¹ studied mice infected with *Salmonella enterica* serovar Typhimurium (*Salmonella typhimurium*). **a**, The bacteria first enter the lumen of the gut, which also contains numerous non-pathogenic commensal bacterial species. However, expression of virulence proteins by *S. typhimurium* causes inflammation of gut tissues, killing off many of the commensal bacteria. By the second day of an *S. typhimurium* infection, *hilD*-mutant bacteria arise that no longer express some virulence proteins; this allows them to grow faster than the wild-type

bacteria. But the authors show that the infecting population is 'bistable' — approximately 60% of the bacteria are phenotypically avirulent through downregulation of virulence-gene expression, and these grow just as fast as the mutants. **b**, When Diard and colleagues manipulated the bacteria so that the infecting population had a lower proportion of phenotypically avirulent cells, they found that the wild-type bacteria were less effective at keeping the mutant bacteria in check. The relative lack of virulent bacteria in this infection led to the inflammation subsiding faster than in the infection shown in **a**.

mice. They predicted that avirulent mutant bacteria would be unable to establish an infection on their own, but that if avirulent mutants arose spontaneously from a population of wild-type bacteria that had already conditioned the host environment, then the mutants would prosper. This was indeed the case. In mice infected with wild-type bacteria, the authors found a small population of defector bacteria by day 2. This population rapidly expanded to dominate the infection (Fig. 1a). Interestingly, all the mutants had a mutation in *hilD*, the 'master regulator' gene for one particular virulence system in *S. typhimurium*¹⁰. Reinforcing their predictions, this mutant type was incapable of establishing infection on its own, but it rapidly dominated during co-infections with wild-type bacteria.

Interestingly, infections that were initiated by a mixture of wild-type and *hilD* mutants were short-lived and were cleared more quickly by the host, demonstrating that mutant defectors undermine the ultimate survival of the population. This is in agreement with a previous observation² that high relatedness in a population is essential for evolutionary stability of cooperative behaviour. Thus, the fitness advantage of accruing mutations that lead to fast-growing avirulent cells is due to social exploitation because this advantage is only apparent in the presence of wild-type cells that can elicit the beneficial inflammatory response from the host.

So how is it that cooperative virulence is maintained when the spontaneous generation of mutant defectors during infection is favoured by selection? Diard and colleagues predicted that a bistable population of genetically identical bacteria that differ in virulence-factor expression might hold a clue. When they modelled a bistable population, the researchers found that

maintaining a high proportion of fast-growing phenotypically avirulent cells kept the rise of mutant defectors in check, whereas lowering the proportion of phenotypically avirulent cells accelerated the emergence of such mutants. To validate this model *in vivo*, the authors manipulated the bistable population distribution by disrupting the protein HilE, a negative regulator of HilD (ref. 11). In these experiments, lowering the proportion of phenotypically avirulent bacteria did make the population more susceptible to mutant defectors, and caused early cessation of inflammation and loss of *S. typhimurium* from the gut (Fig. 1b).

The finding that genetically identical yet phenotypically distinct subpopulations control the rise of avirulent cheats provides fascinating insight into the evolution of virulence. However, questions remain. For any pathogen, the key to evolutionary success lies in its ability to transmit from host to host¹², but it is unclear from the current work whether *S. typhimurium* transmission is actually compromised in the presence of defectors. If it is, this would strengthen the argument¹³ that cooperative virulence is a selective trait. Also unknown is whether the outgrowth of fast-growing phenotypically avirulent cells occurs in response to environmental cues (or even the mutant defectors themselves) or if it is a truly stochastic process. If the outgrowth is programmable, then how an optimal balance of virulent and avirulent populations is achieved and maintained becomes an interesting question.

This work provides a rare *in vivo* analysis of cooperative traits in pathogens and advances the field of social evolution theory for pathogenic microorganisms. The authors' description of *S. typhimurium*'s elegant mechanism to manipulate phenotypic distribution by

bistability should allow for deeper investigation of other mixed populations, such as the presence of non-replicating dormant *Salmonella* in chronic infections. ■

David T. Mulder and Brian K. Coombes are at the Michael G. DeGroote Institute for Infectious Disease Research and the Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8S 4K1, Canada.
e-mail: coombes@mcmaster.ca

1. Diard, M. *et al.* *Nature* **494**, 353–356 (2013).
2. West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. *Nature Rev. Microbiol.* **4**, 597–607 (2006).
3. Dubnau, D. & Losick, R. *Mol. Microbiol.* **61**, 564–572 (2006).
4. Veening, J. W., Smits, W. K. & Kuipers, O. P. *Annu. Rev. Microbiol.* **62**, 193–210 (2008).
5. Stecher, B. *et al.* *PLoS Biol.* **5**, 2177–2189 (2007).
6. Ackermann, M. *et al.* *Nature* **454**, 987–990 (2008).
7. Sturm, A. *et al.* *PLoS Pathog.* **7**, e1002143 (2011).
8. Harrison, F., Browning, L. E., Vos, M. & Buckling, A. *BMC Biol.* **4**, 21 (2006).
9. Rumbaugh, K. P. *et al.* *Curr. Biol.* **19**, 341–345 (2009).
10. Ellermeier, C. D., Ellermeier, J. R. & Schlauch, J. M. *Mol. Microbiol.* **57**, 691–705 (2005).
11. Baxter, M. A., Fahlen, T. F., Wilson, R. L. & Jones, B. D. *Infect. Immun.* **71**, 1295–1305 (2003).
12. Wickham, M. E., Brown, N. F., Boyle, E. C., Coombes, B. K. & Finlay, B. B. *Curr. Biol.* **17**, 783–788 (2007).
13. Raymond, B., West, S. A., Griffin, A. S. & Bonsall, M. B. *Science* **337**, 85–88 (2012).

CORRECTION

In the News & Views article 'Astrophysics: Going supernova' by Alexander Heger (*Nature* **494**, 46–47; 2013), supernova SN 1987A was incorrectly stated as having occurred in the Small Magellanic Cloud galaxy. The correct galaxy is the Large Magellanic Cloud.

FOXO3A directs a protective autophagy program in haematopoietic stem cells

Matthew R. Warr¹, Mikhail Binnewies¹, Johanna Flach¹, Damien Reynaud¹, Trit Garg¹, Ritu Malhotra², Jayanta Debnath² & Emmanuelle Passegue¹

Blood production is ensured by rare, self-renewing haematopoietic stem cells (HSCs). How HSCs accommodate the diverse cellular stresses associated with their life-long activity remains elusive. Here we identify autophagy as an essential mechanism protecting HSCs from metabolic stress. We show that mouse HSCs, in contrast to their short-lived myeloid progeny, robustly induce autophagy after *ex vivo* cytokine withdrawal and *in vivo* calorie restriction. We demonstrate that FOXO3A is critical to maintain a gene expression program that poises HSCs for rapid induction of autophagy upon starvation. Notably, we find that old HSCs retain an intact FOXO3A-driven pro-autophagy gene program, and that ongoing autophagy is needed to mitigate an energy crisis and allow their survival. Our results demonstrate that autophagy is essential for the life-long maintenance of the HSC compartment and for supporting an old, failing blood system.

Understanding how tissue homeostasis is maintained throughout life is a fundamental question in biology, with critical implications for ageing and disease development¹. Every cell of an organism must contend with an array of intrinsic and extrinsic stresses, which are exacerbated with age and place enormous functional demands on each tissue. Self-renewing stem cells must constantly counterbalance the deleterious effects of chronic stress by generating new cells to replace those that become damaged or destroyed. This is particularly critical for the maintenance of tissues with high cellular turnover such as the blood system². Although HSCs efficiently give rise to all mature blood cells in young individuals, they often fail or accumulate damage with age, resulting in the development of serious and often fatal diseases including bone marrow failure syndromes and leukaemias³. Whereas much is known about the mechanisms controlling HSC self-renewal and differentiation activities^{2,4}, less is understood about the pathways used by HSCs to combat cellular stress.

Autophagy and apoptosis are two major stress-response pathways that are often co-regulated but elicit opposite cellular outcomes⁵. Autophagy is a self-catabolic process by which cellular components are sequestered into double-membrane vacuoles, called autophagosomes, and degraded through the lysosomal machinery⁶. In addition to this 'housekeeping' function, autophagy also has an essential 'adaptive' survival role by maintaining nutrient and energy levels during periods of metabolic starvation or stress. In contrast, apoptosis results in the self-elimination of excessively damaged or non-functional cells, which safeguards the overall health of the organism⁷. Common upstream signalling components regulate both apoptosis and autophagy^{8,9}, and numerous experiments reveal that stresses may trigger either process depending on the cellular context. Inhibition of either pathway by both genetic and chemical means can also result in the activation of the other; that is, blocking apoptosis in cells that normally would die can trigger autophagy and survival, whereas blocking autophagy induction in cells that normally would survive rapidly induces apoptosis.

HSCs, like other cells in the body, have the capacity to induce cell death through apoptosis when they become superfluous or irreversibly

damaged¹⁰. Autophagy has also been shown to control essential processes within the blood system, including immune regulation¹¹, lymphocyte survival^{12,13}, erythroid cell maturation^{14,15} and maintenance of fetal HSCs^{16,17}. However, as there are major differences between fetal and adult HSCs, particularly with respect to their cell cycle status¹⁸, it remains to be established whether autophagy has a specific role in protecting quiescent adult HSCs.

Autophagy induction in HSCs

To determine whether haematopoietic stem and progenitor cells can induce autophagy, we isolated HSCs (Lin[−]c-Kit⁺Sca1⁺Flk2[−]CD48[−]CD150⁺) and granulocyte/macrophage progenitors (GMPs; Lin[−]c-Kit⁺Sca1[−]CD34⁺FcγR⁺) from the bone marrow of GFP-LC3 transgenic mice¹⁹ (Supplementary Fig. 1a). We monitored autophagy induction *ex vivo* upon culture with or without cytokines by assaying the formation and turnover of punctate GFP-LC3, indicative of autophagosome formation and autolysosome degradation, respectively²⁰. We first confirmed that cytokine withdrawal induced a metabolic stress as evidenced by decreased glucose uptake in HSCs and GMPs cultured without cytokines (Supplementary Fig. 1b, c). We also treated HSCs and GMPs with the lysosomal inhibitor bafilomycin A (BafA) to block autolysosome destruction and trap the resulting punctate GFP-LC3. Whereas punctate GFP-LC3 was seen in GMPs upon culture with cytokines, autophagosome formation was not further enhanced by cytokine withdrawal (Fig. 1a). In contrast, HSCs displayed no detectable punctate GFP-LC3 when cultured in the presence of cytokines, but exhibited robust puncta formation after cytokine withdrawal (Fig. 1b). We independently validated these observations using transmission electron microscopy²⁰. Although freshly isolated GMPs readily displayed autophagic vacuoles (Supplementary Fig. 2a), we found no signs of autophagy in freshly isolated HSCs and observed autophagic vacuoles only upon starvation (Fig. 1c and Supplementary Fig. 2b). We then used flow cytometry²¹ to measure GFP-LC3 fluorescence levels with or without BafA, which allows for the quantification of autophagy flux in HSCs and GMPs (Supplementary Fig. 3a). As expected, HSCs showed a clear loss in

¹The Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Department of Medicine, Division of Hematology/Oncology, University of California San Francisco, San Francisco, California 94143, USA. ²Department of Pathology, University of California, San Francisco, California 94143, USA.

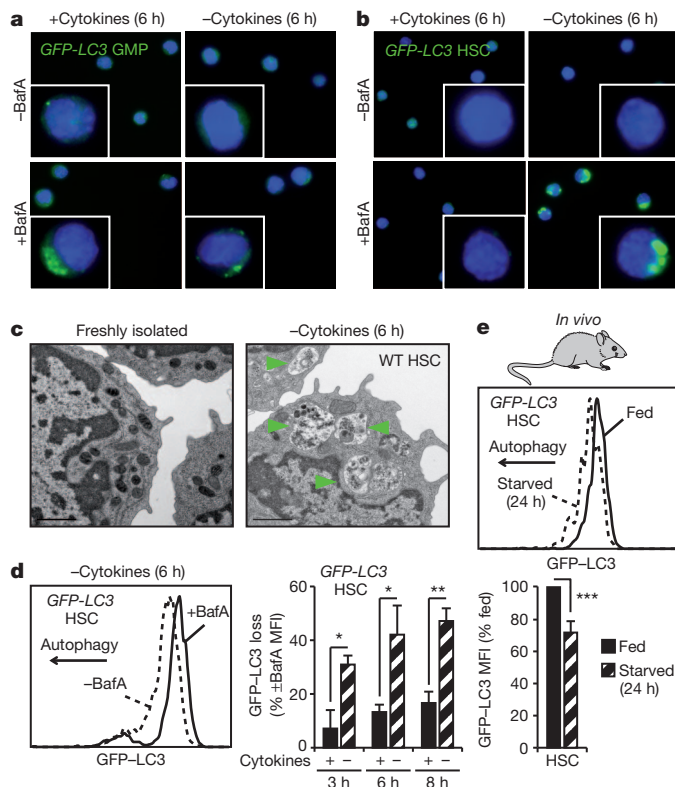


Figure 1 | HSCs induce autophagy after metabolic stress. **a, b**, Fluorescent microscopy of cultured GFP-LC3 GMPs (**a**) and HSCs (**b**). Original magnification, $\times 63$. **c**, Electron microscopy images of wild-type HSCs. Arrowheads indicate autophagic vesicles. Scale bars, 1 μ m. **d**, GFP-LC3 loss in cultured GFP-LC3 HSCs ($n = 3$). Results are expressed as a percentage of GFP-LC3 mean fluorescence intensity (MFI) in -BafA compared to +BafA conditions. **e**, GFP-LC3 levels in HSCs of fed and starved GFP-LC3 mice ($n = 3$). Results are expressed as a percentage of GFP-LC3 mean fluorescence intensity in fed mice. Data are mean \pm s.d. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

GFP-LC3 levels when cultured without cytokines, with minimal changes upon culture with cytokines (Fig. 1d). In contrast, GMPs had a similar loss of GFP-LC3 when cultured with or without cytokines (Supplementary Fig. 3b). Consistent with these analyses, we also found that only starved HSCs, and not starved GMPs, had decreased levels of intracellular p62 protein²⁰ (Supplementary Fig. 3c). Finally, we deprived GFP-LC3 mice of food for 24 h as calorie restriction has been shown to induce autophagy potentially¹⁹. As expected, we observed a significant decrease in GFP-LC3 levels in HSCs, but not myeloid progenitors, hence confirming autophagy induction in the HSC compartment after *in vivo* starvation (Fig. 1e and Supplementary Fig. 3d). Collectively, these results demonstrate that only long-lived HSCs, in contrast to their short-lived myeloid progeny, are able to mount a robust autophagic response upon metabolic stress.

Protective function of autophagy

Tissue culture studies indicate that autophagy provides starved cells with a window of survival, particularly when apoptosis is blocked⁸. To assess whether autophagy allows HSCs to survive conditions of metabolic stress, we cultured HSCs *ex vivo* with or without cytokines and BafA, and monitored apoptosis levels using a cleaved-caspase-3/7-based luminescence assay. Although cytokine withdrawal was sufficient to promote low-level caspase activation, inhibiting autophagy with BafA both accelerated the onset and increased the level of apoptosis in starved HSCs (Fig. 2a). We functionally confirmed these results by transplanting HSCs that were cultured for 12 h with or without cytokines and BafA (Supplementary Fig. 4a). As expected,

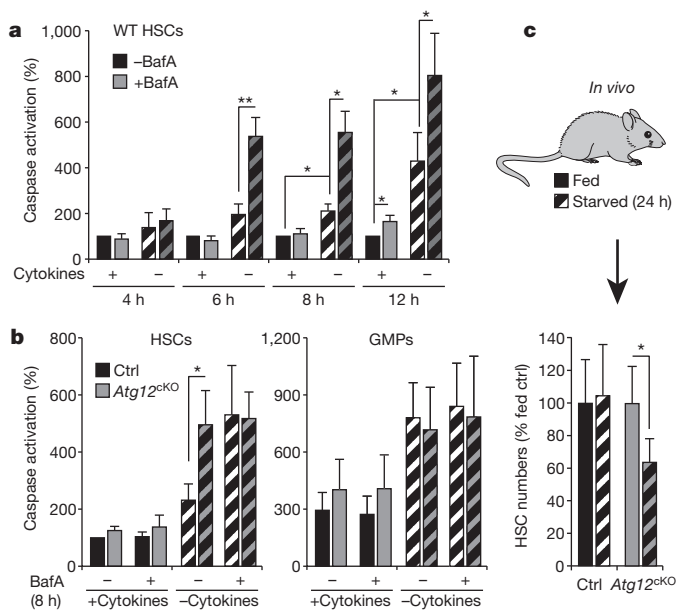


Figure 2 | Autophagy protects HSCs from starvation-induced apoptosis. **a**, Apoptosis levels in cultured wild-type (WT) HSCs ($n = 3-5$). Results are expressed as the percentage of caspase activation compared to +cytokines conditions. **b**, Apoptosis levels in cultured control (ctrl) and *Atg12*^{CKO} HSCs and GMPs ($n = 3$). Results are expressed as the percentage of caspase activation compared to +cytokine control HSCs (100%). **c**, Absolute number of HSCs in fed and starved control and *Atg12*^{CKO} mice ($n = 4-8$ mice per group). Results are expressed as the percentage of HSCs in fed mice. Data are mean \pm s.d. Hatching indicates starved conditions. * $P \leq 0.05$; ** $P \leq 0.01$.

starved HSCs had significantly diminished reconstitution capability, and BafA-mediated blockade of autophagy resulted in a further reduction in the number of engrafting HSCs. To demonstrate that autophagy is directly involved in protecting HSCs, we generated mice with haematopoietic-specific deletion of the essential autophagy machinery component, ATG12 (*Atg12*^{fl/fl}*Mx1-cre*, hereafter called *Atg12*^{CKO})²² (Supplementary Fig. 4b, c). As expected, we observed a significant increase in caspase activation in starved *Atg12*^{CKO} HSCs, which was comparable to that elicited by BafA (Fig. 2b). In contrast, we found no evidence of a protective autophagic response in the apoptosis-prone GMPs²³ as neither *Atg12* deletion nor BafA treatment led to an additional increase in caspase activation above their already elevated basal levels (Fig. 2b). To confirm further that autophagy protects starved HSCs from undergoing apoptosis, we generated mice with haematopoietic-specific deletion of the apoptotic mediators BAK and BAX (*Bak*^{-/-}*Bax*^{fl/fl}*Mx1-cre*, hereafter called *BakBax*^{CKO})²⁴ (Supplementary Fig. 4d). We found that *BakBax*^{CKO} HSCs had no caspase activation in any of the tested culture conditions (Supplementary Fig. 4e), hence validating our *ex vivo* apoptosis assay and demonstrating that the mitochondrial apoptotic pathway is responsible for the death of starved HSCs. We also deprived *Atg12*^{CKO} mice of food for 24 h and observed a significant loss of phenotypic HSCs, which was not observed in control mice (Fig. 2d). Although we were unable to formally document increased apoptosis in starved *Atg12*^{CKO} HSCs (data not shown), our results support an essential role for autophagy in protecting HSCs during calorie restriction *in vivo*. Finally, we confirmed that HSCs, which underwent robust autophagy induction upon either *ex vivo* cytokine withdrawal or *in vivo* calorie restriction, were unchanged in their differentiation properties as measured by colony-forming unit (c.f.u.) assays or lineage reconstitution upon transplantation (Supplementary Fig. 5). Collectively, these data demonstrate that HSCs use autophagy to prevent starvation-induced apoptosis and maintain their functionality.

Regulation of autophagy

To delineate why HSCs are competent for autophagy induction, whereas GMPs are not, we first measured the activation status of mammalian target of rapamycin complex 1 (mTORC1), the principal pathway restricting autophagy induction⁶. However, we did not observe differences in mTORC1 activation levels between HSCs and myeloid progenitors using intracellular flow cytometry staining for either phosphorylated mTOR (phospho-mTOR), or its direct target phospho-S6 (Supplementary Fig. 6a). We also did not find any major differences in phospho-S6 levels in HSCs and GMPs cultured for 6 h with cytokines, and confirmed that both were sensitive to rapamycin-mediated mTORC1 inhibition (Supplementary Fig. 6b). These results indicate that the inability of GMPs to induce autophagy upon metabolic stress is not secondary to elevated basal levels of mTORC1 activation. We then used a commercial autophagy PCR array to interrogate the status of the autophagy machinery in HSCs, GMPs and terminally differentiated granulocytes (Gr1⁺Mac1⁺) (Supplementary Fig. 7a). Interestingly, granulocytes had a clear decrease in the expression of nearly all autophagy genes when compared to GMPs, indicating that mature myeloid cells lack the autophagy machinery needed to induce autophagic flux (Supplementary Fig. 7b and Supplementary Table 1). Moreover, we found that expression of the core autophagy machinery components was largely unchanged between HSCs and GMPs; nevertheless, many key upstream regulators and all the Atg8 homologues, which are turned over during autophagic proteolysis⁶, were specifically upregulated in HSCs (Fig. 3a and Supplementary Table 2). We extracted a pro-autophagic gene expression signature, and directly confirmed by quantitative RT-PCR the increased expression of these genes in HSCs compared to GMPs (Fig. 3b). Collectively, these results support the idea that long-lived HSCs are molecularly wired to mount a protective survival autophagy response upon metabolic stress, whereas their short-lived myeloid progeny are not.

FOXO3A-driven pro-autophagy gene program

We next searched for specific transcriptional regulators that could maintain high expression of pro-autophagy genes in HSCs. We identified FOXO3A and p53 as two putative candidates (Supplementary Fig. 8) that can both trigger autophagy and apoptosis through induction of a largely overlapping set of targets^{25,26}. We already showed that HSCs and GMPs expressed p53 at similar levels²³, and confirmed here that only HSCs had transcriptionally active nuclear FOXO3A (Supplementary Fig. 9a). In addition, we found higher expression levels of FOXO3A, and other family members, in HSCs compared to GMPs (Supplementary Fig. 9b). Notably, FOXO3A deletion led to a marked downregulation of the pro-autophagic program, with six of our identified targets showing significantly reduced expression in *Foxo3a*^{-/-} HSCs (Fig. 3c). We then crossed *Foxo3a*^{-/-} mice with GFP-LC3 transgenic mice to assess how lack of FOXO3A affected the ability of HSCs to mount an adaptive autophagy response. We observed a nearly twofold increase in the basal levels of GFP-LC3 in *Foxo3a*^{-/-} GFP-LC3 HSCs, consistent with constitutively lower levels of autophagic flux compared to control HSCs (Supplementary Fig. 9c). Moreover, we found a marked reduction in autophagy induction in *Foxo3a*^{-/-} GFP-LC3 HSCs, as demonstrated by the delayed loss of GFP-LC3 compared to control HSCs after 3 h and 6 h culture without cytokines (Fig. 3d and Supplementary Fig. 9d). In contrast, we did not find any significant changes in the expression level of pro-autophagic genes in *Trp53*^{-/-} HSCs (Supplementary Fig. 10a), nor in the basal GFP-LC3 levels or loss of GFP-LC3 upon cytokine withdrawal in *Trp53*^{-/-} GFP-LC3 HSCs (Supplementary Fig. 10b, c). Furthermore, we generated double-knockout *Foxo3a*^{-/-} *p53*^{-/-} GFP-LC3 mice and did not observe any additional increase in the basal GFP-LC3 levels or further delay in autophagy induction after cytokine withdrawal compared to *Foxo3a*^{-/-} GFP-LC3 HSCs (Supplementary Fig. 11a, b). These results rule out a major role for the p53 pathway for autophagy induction in HSCs, and

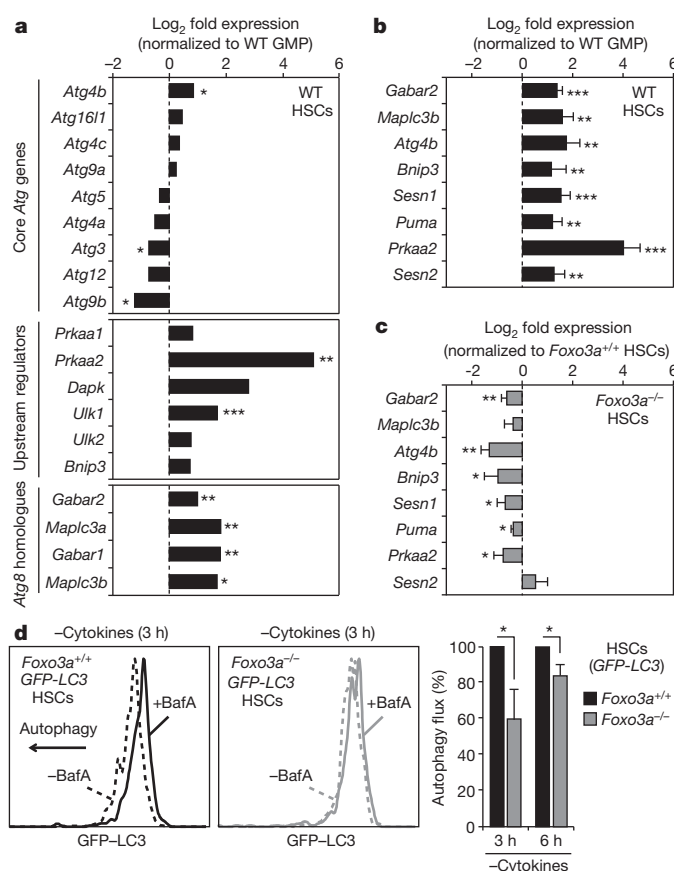


Figure 3 | FOXO3A poises HSCs for rapid autophagy induction. **a**, Status of the autophagy machinery in wild-type HSCs ($n = 3$). Results are expressed as log₂ fold expression compared to wild-type GMPs (set to 0). **b**, **c**, qRT-PCR analyses of pro-autophagic genes in wild-type (**b**) and *Foxo3a*^{-/-} (**c**) HSCs ($n = 3$). Results are expressed as log₂ fold expression compared to wild-type GMPs (**b**) or *Foxo3a*^{+/+} HSCs (**c**) (set to 0). **d**, Autophagy flux in cultured *Foxo3a*^{-/-} GFP-LC3 HSCs ($n = 3$). Results are expressed as the percentage of GFP-LC3 mean fluorescence intensity in -BafA versus +BafA conditions, and normalized to *Foxo3a*^{+/+} GFP-LC3 HSCs. Data are mean \pm s.d. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

suggest that functional compensation by other FOXO family members²⁷ may account for the delayed autophagic response occurring in the absence of FOXO3A. Collectively, these experiments demonstrate that FOXO3A maintains a pro-autophagy gene expression program that specifically poises HSCs to rapidly mount a protective autophagic response upon metabolic stress.

Autophagy in old HSCs

Decreased capacity to undergo autophagy has been proposed to be associated with ageing^{1,28}, and functional deficiencies in old HSCs underlie many of the blood pathologies associated with old age³. Thus, we investigated whether HSCs isolated from old mice (≥ 24 months of age) displayed reduced autophagic potential compared to HSCs isolated from young mice (6–12 weeks of age) (Supplementary Fig. 12a). Unexpectedly, electron microscopy analyses of freshly isolated old, wild-type HSCs revealed the presence of numerous autophagic vacuoles, which were absent in young HSCs (Fig. 4). In addition, old GFP-LC3 HSCs, but not myeloid progenitors, had decreased basal levels of GFP-LC3, and punctate GFP-LC3 was detectable even in the presence of cytokines in BafA-treated old GFP-LC3 HSCs (Supplementary Fig. 12b, 13a). Furthermore, like young GFP-LC3 HSCs, old GFP-LC3 HSCs still underwent robust autophagic flux upon 6-h cytokine withdrawal *ex vivo* (Supplementary Fig. 13b). It was previously shown that old HSCs had elevated levels of activated mTORC1 (ref. 29) that

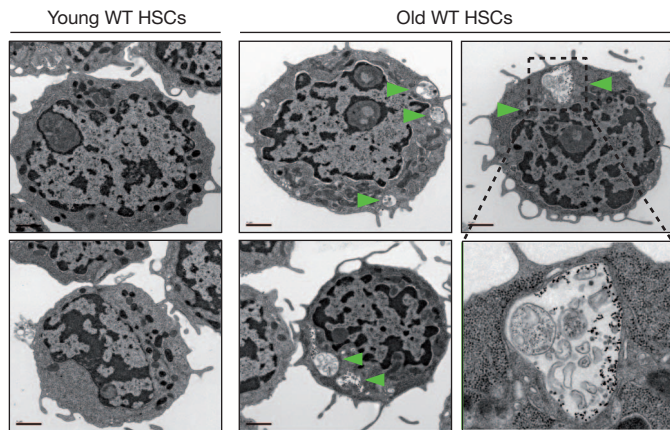


Figure 4 | Ongoing autophagy in old HSCs. Electron microscopy images of young and old HSCs. Arrowheads indicate autophagic vesicles. Scale bars, 1 μ m.

should, in principle, prevent such induction of autophagy. However, our analyses of phospho-mTOR and phospho-S6 levels revealed no significant differences in mTORC1 activity between young and old HSCs (Supplementary Fig. 14a). Collectively, these results indicate that old HSCs have higher basal autophagic flux than young HSCs, but retain their ability to induce autophagy upon metabolic stress.

To determine whether old HSCs are similarly poised to induce autophagy as their younger counterparts, we next investigated the expression status of their autophagy machinery (Supplementary Fig. 14b and Supplementary Table 3). We found that old HSCs had unchanged expression of all major components of the core autophagy machinery, upstream regulators and *Atg8* homologues compared to young HSCs. We also confirmed that old HSCs still had nuclear FOXO3A expression (Supplementary Fig. 14c) and maintained expression of direct FOXO3A pro-autophagy targets (Fig. 5a). Although expression of the p53 co-regulated *Puma* (also called *Bbc3*), *Prkaa2* and *Sesn2* targets was significantly reduced in old HSCs, it did not reflect a global decrease in p53 activity, as expression of other known p53 targets was either unchanged or even increased (Supplementary Fig. 14d). Although old HSCs displayed overall lower levels of apoptosis, treatment with BafA significantly increased caspase activation to a level that eventually reached the apoptotic response elicited in starved young HSCs, albeit at a later time point (Fig. 5b and Supplementary Fig. 15a). Altogether, these results demonstrate that autophagy induction provides a similar survival benefit to young and old HSCs.

To determine the function of basal autophagy flux in old HSCs, we then plated young and old HSCs in methylcellulose containing BafA (Supplementary Fig. 15b). Remarkably, BafA-mediated blockade of autophagy resulted in the specific loss of cloning efficiency of old HSCs, with no effect on young HSCs, which indicates that ongoing autophagy is actually required for the survival of old HSCs. Because a major function of autophagy is to mitigate oxidative stress via the removal of toxic proteins and damaged mitochondria⁶, we used the fluorescence-based dichlorofluorescein diacetate (DCFDA) probe to measure reactive oxygen species (ROS) levels. However, we found similar ROS levels at steady state and an equivalent increase upon BafA-mediated blockade of autophagy in both young and old HSCs (Supplementary Fig. 15c). We next used the fluorescently labelled deoxyglucose analogue 2-NBD glucose to directly compare the ability of young and old HSCs to uptake glucose in culture conditions with cytokines (Fig. 5c and Supplementary Fig. 15d). Notably, we found that old HSCs had significantly attenuated 2-NBD glucose uptake compared to young HSCs, suggesting that impaired nutrient uptake could be a major driver for the heightened autophagy flux in old HSCs. This would be consistent with the recently reported decrease in cytokine availability in the old bone marrow microenvironment³⁰. To directly assess whether old HSCs need autophagy to preserve

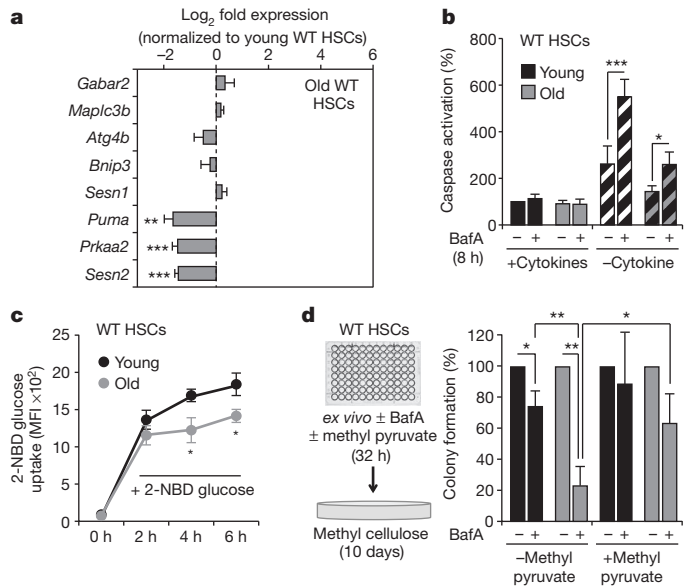


Figure 5 | Ongoing autophagy is essential for the continued survival of old HSCs. **a**, qRT-PCR analyses of pro-autophagic genes in old HSCs ($n = 3-5$). Results are expressed as \log_2 fold expression compared to young HSCs (set to 0). **b**, Apoptosis levels in cultured young and old HSCs ($n = 3$). Results are expressed as the percentage of caspase activation compared to +cytokine young HSCs. Hatching indicates -cytokine conditions. **c**, 2-NBD glucose uptake in cultured young and old HSCs ($n = 3$). **d**, Percentage colony formation in young and old wild-type HSCs cultured for 32 h with or without BafA and methyl pyruvate ($n = 3$). Colonies were counted at day 10 and normalized to -BafA conditions. Data are mean \pm s.d. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

energy levels, we then co-cultured old and young HSCs for 32 h with or without BafA and methyl pyruvate, a cell-permeable form of pyruvate that alleviates the need for nutrient uptake to maintain ATP production³¹. Remarkably, incubation with methyl pyruvate significantly rescued the decreased plating efficiency of old HSCs observed upon BafA-mediated blockade of autophagy (Fig. 5d). Collectively, these results demonstrate that old HSCs do not have reduced autophagy potential but, in fact, rely on autophagy for their continued survival in response to reduced nutrient uptake and energy production.

Discussion

We show that autophagy is essential to protect adult HSCs from metabolic stress and, probably, permits the HSC compartment to withstand the growth factor fluctuations and nutrient deprivations that naturally occur in the bone marrow microenvironment, especially after haematopoietic injury³². We establish that HSCs are uniquely wired to mount an autophagy-mediated cytoprotective response due to the expression of a FOXO3A-driven pro-autophagy gene expression program that poises them for rapid induction of autophagy after metabolic stress (Supplementary Fig. 16). Our results are consistent with the haematopoietic phenotype of *Foxo3a*^{-/-} mice³³, which have a diminished HSC pool size later in life, and indicate that in the absence of robust autophagic capacity, HSCs are no longer fully protected *in vivo*.

Our work also challenges the prevailing view in the ageing field that impaired autophagy is one of the principal determinants of cellular ageing²⁸. Instead, we found that the regulatory pathways that poise HSCs for rapid induction of autophagy remain intact during physiological ageing, and that ongoing autophagy is essential for the continued survival of old HSCs. In fact, old HSCs have been shown to accumulate DNA damage and to often lose their functionality, which are believed to be at the root of the elevated incidence of blood disorders in the elderly³. Here we show that old HSCs also have decreased capacity for nutrient uptake, potentially due to the untoward effects of

an aged bone marrow environment³⁰, and thus rely on ongoing autophagy to maintain their energy levels and survive in these cytokine-poor conditions (Supplementary Fig. 16). This finding demonstrates that autophagy directly protects the old, albeit inefficient, blood system. It will now be interesting to determine whether autophagy also indirectly contributes to the development of age-related blood diseases by allowing the survival of damaged, dysfunctional or transformed old HSCs that would routinely be eliminated via apoptosis in younger animals.

METHODS SUMMARY

Young mice were 6–12 weeks of age and old mice were ≥ 24 months of age. All mouse strains, except for *Atg12^{fl/fl}*, and protocols for cell isolation and transplantation procedures were described previously. Cells were cultured in StemPro-34 or methylcellulose medium containing or not cytokines (IL-3, GM-CSF, SCF, IL-11, Flt-3L, Tpo, Epo), bafilomycin A (BafA, 5 nM) and methyl pyruvate (10 mM). Apoptosis assays were performed in 384-well plates using the Caspase-Glo 3/7 Promega protocol. GFP-LC3 fluorescent microscopy and immunostainings were performed on cells directly sorted onto poly-lysine-coated slides and fixed in 4% PFA. Intracellular phospho-staining, flow cytometry analyses of DCFDA-based ROS production and 2-NBD glucose uptake, and RNA isolation followed by qRT-PCR analyses were performed as described previously. Mouse Autophagy RT² Profiler PCR Arrays were purchased from SABiosciences and used as recommended by the manufacturer. All data were expressed as mean \pm standard deviation (s.d.). *P* values were generated using two-tailed student's *t*-test.

Full Methods and any associated references are available in the online version of the paper.

Received 8 April 2012; accepted 8 January 2013.

Published online 6 February 2013.

- Rando, T. A. Stem cells, ageing and the quest for immortality. *Nature* **441**, 1080–1086 (2006).
- Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
- Rossi, D. J., Jamieson, C. H. M. & Weissman, I. L. Stems cells and the pathways to aging and cancer. *Cell* **132**, 681–696 (2008).
- Warr, M. R., Pietras, E. M. & Passequé, E. Mechanisms controlling hematopoietic stem cell functions during normal hematopoiesis and hematological malignancies. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 681–701 (2011).
- Ferraro, E. & Cecconi, F. Autophagic and apoptotic response to stress signals in mammalian cells. *Arch. Biochem. Biophys.* **462**, 210–219 (2007).
- He, C. & Klionsky, D. J. Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.* **43**, 67–93 (2009).
- Danial, N. N. & Korsmeyer, S. J. Cell death: critical control points. *Cell* **116**, 205–219 (2004).
- Maiuri, M. C., Zalckvar, E., Kimchi, A. & Kroemer, G. Self-eating and self-killing: crosstalk between autophagy and apoptosis. *Nature Rev. Mol. Cell Biol.* **8**, 741–752 (2007).
- Debnath, J., Baehrecke, E. H. & Kroemer, G. Does autophagy contribute to cell death? *Autophagy* **1**, 66–74 (2005).
- Oguro, H. & Iwana, A. Life and death in hematopoietic stem cells. *Curr. Opin. Immunol.* **19**, 503–509 (2007).
- Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
- Pua, H. H., Dzhagalov, I., Chuck, M., Mizushima, N. & He, Y. W. A critical role for the autophagy gene *Atg5* in T cell survival and proliferation. *J. Exp. Med.* **204**, 25–31 (2007).
- Miller, B. C. *et al.* The autophagy gene *ATG5* plays an essential role in B lymphocyte development. *Autophagy* **4**, 309–314 (2008).
- Kundu, M. *et al.* Ulk1 plays a critical role in the autophagic clearance of mitochondria and ribosomes during reticulocyte maturation. *Blood* **112**, 1493–1502 (2008).
- Mortensen, M. *et al.* Loss of autophagy in erythroid cells leads to defective removal of mitochondria and severe anemia *in vivo*. *Proc. Natl Acad. Sci. USA* **107**, 832–837 (2010).
- Liu, F. *et al.* FIP200 is required for the cell-autonomous maintenance of fetal hematopoietic stem cells. *Blood* **116**, 4806–4814 (2010).
- Mortensen, M. *et al.* The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. *J. Exp. Med.* **208**, 455–467 (2011).
- Pietras, E. M., Warr, M. R. & Passequé, E. Cell cycle regulation in hematopoietic stem cells. *J. Cell Biol.* **195**, 709–720 (2011).
- Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T. & Ohsumi, Y. *In vivo* analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. *Mol. Biol. Cell* **15**, 1101–1111 (2004).
- Mizushima, N., Yoshimori, T. & Levine, B. Methods in mammalian autophagy research. *Cell* **140**, 313–326 (2010).
- Shvets, E., Fass, E. & Elazar, Z. Utilizing flow cytometry to monitor autophagy in living mammalian cells. *Autophagy* **4**, 621–628 (2008).
- Radoshevich, L. *et al.* ATG12 conjugation to ATG3 regulates mitochondrial homeostasis and cell death. *Cell* **142**, 590–600 (2010).
- Mohrin, M. *et al.* Hematopoietic stem cell quiescence promotes error prone DNA repair and mutagenesis. *Cell Stem Cell* **7**, 174–185 (2010).
- Lindsten, T. *et al.* The combined functions of proapoptotic Bcl-2 family members *bak* and *bax* are essential for normal development of multiple tissues. *Mol. Cell* **6**, 1389–1399 (2000).
- Mammucari, C. *et al.* FoxO3 controls autophagy in skeletal muscle *in vivo*. *Cell Metab.* **6**, 458–471 (2007).
- Maiuri, M. C. *et al.* Autophagy regulation by p53. *Curr. Opin. Cell Biol.* **22**, 181–185 (2010).
- Tothova, Z. *et al.* FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007).
- Rubinsztein, D. C., Marino, G. & Kroemer, G. Autophagy and aging. *Cell* **146**, 682–695 (2011).
- Chen, C., Liu, Y. & Zheng, P. mTOR regulation and therapeutic rejuvenation of aging hematopoietic stem cells. *Sci. Signal.* **2**, ra75 (2009).
- Ergen, A. V., Boles, N. C. & Goodell, M. A. Rantes/Ccl5 influences hematopoietic stem cell subtypes and causes myeloid skewing. *Blood* **119**, 2500–2509 (2012).
- Lum, J. J. *et al.* Growth factor regulation of autophagy and cell survival in the absence of apoptosis. *Cell* **120**, 237–248 (2005).
- Wilson, A., Laurenti, E. & Trumpp, A. Balancing dormant and self-renewing hematopoietic stem cells. *Curr. Opin. Genet. Dev.* **19**, 461–468 (2009).
- Miyamoto, K. *et al.* Foxo3a is essential for maintenance of the hematopoietic stem cell pool. *Cell Stem Cell* **1**, 101–112 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Brunet for *Foxo3a^{-/-}* mice, S. Oakes for *Bak^{-/-}* *Bax^{fl/fl}* mice, N. Mizushima for *GFP-LC3* transgenic mice, J. Wong for electron microscopy analyses, M. Kissner for management of our Flow Cytometry Core Facility, and all members of the Passequé, particularly E. Pietras, and Debnath laboratories for critical insights and suggestions. M.R.W. is an LLS Special Fellow. This work was supported by a Culpepper Scholar Award (Partnership for Cures) and NIH grant CA126792 to J.D., and a CIRM New Faculty Award and NIH grant HL092471 to E.P.

Author Contributions M.R.W., M.B., J.F., D.R. and T.G. performed all of the experiments. R.M. and J.D. generated the *Atg12^{fl/fl}* mice. M.R.W., M.B., J.D. and E.P. designed the experiments and interpreted the results. M.R.W. and E.P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.P. (PassequéE@stemcell.ucsf.edu).

METHODS

Mice. Young (6–12 weeks of age) and old (20–24 months of age) wild-type C57BL/6 mice were either bred and aged in house, or purchased from the NIA aged rodent colonies. *Foxo3a*^{-/-} (FVB/N)³⁴, *Trp53*^{-/-} (FVB/N)³⁵, *Bak*^{-/-} *Bax*^{fl/fl} (C57BL/6)³⁶, *Mx1-cre* (C57BL/6)³⁷ and *GFP-LC3* (C57BL/6)¹⁹ mice have been described previously. *Trp53*^{-/-} *GFP-LC3* and *Foxo3a*^{-/-} *GFP-LC3* mice were generated on a mixed FVB/N and C57BL/6 background. The description of *Atg12*^{fl/fl} mice will be provided in a separate publication (R.M. and J.D., manuscript in preparation). For *Mx1-cre*-mediated deletion, 4-week-old *Atg12*^{fl/fl} *Mx1-cre* and *Bak*^{-/-} *Bax*^{fl/fl} *Mx1-cre* mice were injected intraperitoneally three times at 2-day intervals with 250 µg of poly(I/C) (GE Healthcare) in 200 µl PBS. Poly(I/C)-injected *Atg12*^{fl/fl} and *Bak*^{-/-} *Bax*^{fl/fl} mice were used as controls. For transplantation experiments, 8–12-week-old CD45.1 C57BL/6 recipient mice were lethally irradiated (11 Gy, delivered in split dose 3 h apart) and injected retro-orbitally with 500 purified CD45.2 HSCs (with or without *ex vivo* culture) delivered with 300,000 Sca-1-depleted helper CD45.1 bone marrow cells. Transplanted mice were given antibiotic-containing water for at least 4 weeks after irradiation. For *in vivo* starvation experiments, mice were deprived of food for 24 h while keeping free access to water. All animal experiments were performed in accordance with UCSF IACUC approved protocols.

Flow cytometry. Granulocytes (Gr-1⁺ Mac-1⁺) were isolated from unfractionated bone marrow using Gr-1-PB (eBioscience, 57-5931-82) and Mac-1-PE-Cy7 (eBioscience, 25-0112-82) antibodies. Enrichment and cell staining procedures for isolating HSCs (Lin⁻ c-Kit⁺ Sca-1⁺ Flk2⁻ CD48⁻ CD150⁺) and GMPs (Lin⁻ c-Kit⁺ Sca-1⁻ CD34⁺ FcγR⁺) from c-Kit-enriched bone marrow were performed as described^{23,38}. For wild-type mice, HSC and GMPs were stained together using un-conjugated lineage antibodies (Gr-1, Mac1, B220, CD3, CD4, CD5, CD8, Ter-119, all prepared by the UCSF Hybridoma Core Facility) and goat anti-rat-PE-Cy5 (Invitrogen, A10691), c-Kit-APC-eFluor 780 (eBioscience, 47-1171-82), Sca-1-PB (BioLegend, 108120), Flk2-Bio (eBioscience, 13-1351-82), CD48-Alexa Fluor 647 (BioLegend, 103416), CD150-PE (BioLegend, 115904), FcγR-PerCP-eFluor 710 (eBioscience, 46-0161-82), CD34-FITC (eBioscience, 11-0341-85) and SA-PE-Cy7 (eBioscience, 25-4317-82) antibodies. For *GFP-LC3* mice, HSCs and GMPs were stained separately omitting FcγR and CD34 for HSC isolation, and using CD34-Bio (BioLegend, 119304) and omitting Flk2 for GMP isolation. For monitoring autophagy induction in *GFP-LC3* mice *in vivo*, directly conjugated lineage CD3-PE-Cy5 (eBioscience, 15-0031), CD4-PE-Cy5 (eBioscience, 15-0041-82), CD5-PE-Cy5 (BioLegend, 100610), CD8-PE-Cy5 (eBioscience, 15-0081-82), Mac1-PE-Cy5 (eBioscience, 15-0112-82), Gr1-PE-Cy5 (eBioscience, 15-5931-82), B220-PE-Cy5 (eBioscience, 15-0452-82), Ter119-PE-Cy5 (eBioscience, 15-5921-82) and Sca1-PE-Cy7 (BioLegend, 108113) antibodies were used while Flk2 was omitted from the staining. Stained cells were re-suspended for final analysis in Hanks Buffered Salt Solution (HBSS) with 2% heat-inactivated fetal calf serum (FCS) and 1 µg ml⁻¹ propidium iodide for dead cell exclusion. For intracellular staining²⁹, Sca1-PE-Cy7 was used and Flk2, FcγR and CD34 were omitted from the staining. Stained cells were washed in PBS, fixed in Cytofix/Cytoperm buffer (BD Biosciences) for 2 h at 4 °C, washed in PermWash (BD Biosciences), permeabilized with CytoPerm Plus (BD Biosciences) for 10 min at room temperature, re-fixed in Cytofix/Cytoperm buffer for 10 min at 4 °C, washed in PermWash and incubated in PermWash overnight at 4 °C with the following primary antibodies: rabbit anti-phospho-mTOR (Cell Signaling, 5536), rabbit anti-phospho-S6-A488 (Cell Signaling, 4854), guinea-pig anti-p62 (Progen, GP62-C), rabbit IgG control isotype (Cell Signaling, 3900S) or rabbit IgG-A488 conjugate (Cell Signaling, 2975). Cells were then washed in PermWash, eventually incubated with anti-rabbit A488 (Invitrogen, A21206) or anti-guinea-pig A488 (Invitrogen, A11073) secondary antibodies in PermWash for 2 h at 4 °C, and finally re-suspended in PBS/2% FCS for analysis. Peripheral blood was obtained from retro-orbital bleeding and collected in 4 ml of ACK (150 mM NH₄Cl/10 mM KHCO₃) containing 10 mM EDTA. Donor and recipient cells were distinguished by expression of different allelic forms of CD45 (CD45.1-FITC, eBioscience, 11-0454-85 versus CD45.2-PE, eBioscience, 12-0453) and multilineage reconstitution was assessed by expression of myeloid (Mac-1-PE-cy7, eBioscience, 25-0112-82; Gr-1-eFluor450, eBioscience, 48-5931-82), B cells (B220-APC-eFluor780, eBioscience, 47-0452-82) and T cells (CD3-A647, eBioscience, 51-0032-82). Cell isolation was performed on a FACS ARIAII (Becton Dickinson), using double sorting to ensure maximum purity, and cell analysis on a FACS LSRII (Becton Dickinson).

Cell culture. Cells were cultured *ex vivo* in StemPro-34 medium (Invitrogen) supplemented with L-glutamine (2 mM), penicillin (50 U ml⁻¹)/streptomycin (50 µg ml⁻¹), and containing or not the following cytokines (all from PeproTech): IL-3 (10 ng ml⁻¹), GM-CSF (10 ng ml⁻¹), SCF (25 ng ml⁻¹), IL-11 (25 ng ml⁻¹), Flt-3L (25 ng ml⁻¹), Tpo (25 ng ml⁻¹) and Epo (4 U ml⁻¹). Bafilomycin A (BafA, 5 nM, Sigma), methyl pyruvate (10 mM, Sigma), rapamycin (Rapa, 20 µM, EMD Millipore) or appropriate vehicles (DMSO for BafA and

Rapa) were added to the media as indicated. All cultures were performed at 37 °C in a 5% CO₂ water jacket incubator (Thermo Scientific). For apoptosis assays, HSCs and GMPs were sorted directly into 384-well solid white luminiscence plates (400 cells per well, triplicate wells per conditions) containing 40 µl of media (with or without cytokine and BafA). At given time points, 40 µl of caspase-Glo 3/7 (Promega) was added to each well and the plates were shaken at 300 r.p.m. for 30 s, incubated for 45 min at room temperature and read on a luminometer (Synergy2, BioTek) to obtain relative luciferase units (RLU). For each experimental condition, the percentage of caspase activation was calculated as follows using the average RLU obtained from triplicate measurements: (test sample RLU/(HSC + cytokine control sample RLU)) × 100. For colony formation assays, HSCs (100 cells per ml per 3-cm dish, 1 cell per 100 µl per well of 96-well plates) were cultured in methylcellulose (Stem Cell Technologies, M3231) supplemented with L-glutamine, penicillin/streptomycin and all the cytokines described above (with or without BafA when indicated). Colonies (>30 cells) were counted after 10 days of culture. For methyl pyruvate rescue experiments, HSCs (100 cells per well, triplicate wells per conditions) were cultured for 32 h in media containing cytokines (with or without BafA and methyl pyruvate) and then each well was plated in one 3-cm dish of methylcellulose devoid of BafA and methyl pyruvate. For measurement of ROS levels²⁷, HSCs (2,000 cells per well, triplicate wells per conditions) were cultured for 24 h in media (with or without cytokines and BafA). Cells were then washed with HBSS, incubated 15 min at 37 °C with dichlorofluorescein diacetate (DCFDA, 10 µM, Invitrogen) or anhydrous DMSO vehicle, washed with HBSS and directly analysed by flow cytometry for DCFDA fluorescence in the FITC channel. For measurement of nutrient uptake³⁹, HSCs (2,000 cells per well, triplicate wells per conditions) were cultured for the indicated times in cytokine-containing media with or without 2-(*n*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino)-2-deoxyglucose (2-NBD glucose, 100 µM, Invitrogen). Cells were then washed once in HBSS and analysed by flow cytometry for 2-NBD glucose fluorescence in the FITC channel.

Gene expression analyses. Specific pre-tested qRT-PCR primers (*Gabar2*, *Maplc3b*, *Atg4b*, *Bnip3*, *Sesn1*, *Prkaa2*, *Sesn2*, *Dram*) and mouse Autophagy RT² Profiler PCR arrays were purchased from SABiosciences. Other qRT-PCR primer sequences are available upon request. Total RNA was isolated from 10,000–20,000 cells sorted directly into TRIzol-LS (Invitrogen) according to the manufacturer's protocol. For qRT-PCR, RNA was treated with DNase I and reverse-transcribed using SuperScript III kit (Invitrogen). Runs were performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems) using the SYBR Green reagents (Applied Biosystems) and the cDNA equivalent of 200 cells per reaction. Values were normalized to β-actin expression. For PCR array, RNA was further purified using Arcturus PicoPure RNA Isolation kit (Applied Biosystems), which included an on-column DNase I digestion. RNA concentration was measured on a bioanalyser chip and 5 ng of starting RNA was used per sample. Arrays were performed according to the manufacturer's protocol (SABiosciences), with all cDNA samples undergoing pre-amplification (12 cycles) for pathway specific genes. Runs were performed on the same 7900HT Fast Real-Time PCR System and data were analysed using the SABiosciences software. Values were normalized to β-actin expression.

Punctate GFP-LC3 and immunofluorescence staining. Cells were sorted directly onto poly-lysine-coated slides (500–2,000 cells per slide), and fixed in PBS/4% PFA for 10 min at room temperature and treated as follow. For imaging punctate GFP-LC3, slides were directly mounted using ProLong Gold antifade reagent containing 1 µg ml⁻¹ DAPI (Life Technologies) and imaged on a Zeiss Axiovert 200 microscope (63×, 1.4NA objective) driven by Metamorph (v6.0) software (Molecular Devices). For FOXO3A immunofluorescence staining, slides were permeabilized in PBS/0.15% Triton X-100 for 2 min at room temperature and blocked in PBS/1% BSA overnight at 4 °C. Slides were then incubated in PBS/1% BSA with a rabbit anti-FOXO3A (Millipore, 07-1719) for 1 h at 37 °C, washed 3 times in PBS and incubated in PBS/1% BSA with an A594-conjugated goat anti-rabbit secondary antibody (Life Technologies, A-11037) for 1 h at 37 °C. Slides were then washed 3 times in PBS and mounted using VectaShield (Vector Laboratories) containing 1 µg ml⁻¹ DAPI. Cells were imaged on a SP5 Leica Upright Confocal Microscope (×100 objective) and images were processed using Velocity software (v.4.4, Improvision).

Electron microscopy. Cells (50,000–100,000 per condition) were pelleted for 5 min at 4 °C at 600g, fixed on ice for 30 min in 0.1 M NaCacodylate, pH 7.4, containing 2% glutaraldehyde and 1% PFA, and pelleted at 3,000g for 10 min at 4 °C. Samples were then submitted to the Gladstone Institute (UCSF) Electron Microscopy Core Facility for standard transmission electron microscopy ultrastructural analyses.

Statistics. All data were expressed as mean ± standard deviation (s.d.). *P* values were generated using two-tailed student's *t*-test (**P* ≤ 0.05, ***P* ≤ 0.01, ****P* ≤ 0.001).

34. Castrillon, D. H., Miao, L., Kollipara, R., Horner, J. W. & DePinho, R. A. Suppression of ovarian follicle activation in mice by the transcription factor FoxO3a. *Science* **301**, 215–218 (2003).
35. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
36. Takeuchi, O. *et al.* Essential role of BAX, BAK in B cell homeostasis and prevention of autoimmune disease. *Proc. Natl Acad. Sci. USA* **102**, 11272–11277 (2005).
37. Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
38. Santaguida, M. *et al.* JunB protects against myeloid malignancies by limiting hematopoietic stem cell proliferation and differentiation without affecting self-renewal. *Cancer Cell* **15**, 341–352 (2009).
39. Lock, R. *et al.* Autophagy facilitates glycolysis during Ras-mediated oncogenic transformation. *Mol. Biol. Cell* **22**, 165–178 (2011).

A black-hole mass measurement from molecular gas kinematics in NGC4526

Timothy A. Davis¹, Martin Bureau², Michele Cappellari², Marc Sarzi³ & Leo Blitz⁴

The masses of the supermassive black holes found in galaxy bulges are correlated with a multitude of galaxy properties^{1,2}, leading to suggestions that galaxies and black holes may evolve together³. The number of reliably measured black-hole masses is small, and the number of methods for measuring them is limited⁴, holding back attempts to understand this co-evolution. Directly measuring black-hole masses is currently possible with stellar kinematics (in early-type galaxies), ionized-gas kinematics (in some spiral and early-type galaxies^{5–7}) and in rare objects that have central maser emission⁸. Here we report that by modelling the effect of a black hole on the kinematics of molecular gas it is possible to fit interferometric observations of CO emission and thereby accurately estimate black-hole masses. We study the dynamics of the gas in the early-type galaxy NGC 4526, and obtain a best fit that requires the presence of a central dark object of $4.5^{+4.2}_{-3.1} \times 10^8$ solar masses (3σ confidence limit). With the next-generation millimetre-wavelength interferometers these observations could be reproduced in galaxies out to 75 megaparsecs in less than 5 hours of observing time. The use of molecular gas as a kinematic tracer should thus allow one to estimate black-hole masses in hundreds of galaxies in the local Universe, many more than are accessible with current techniques.

For this study we targeted the fast-rotating early-type galaxy (ETG) NGC 4526. This object has a stellar velocity dispersion within one effective radius⁹ $\sigma_e = 222 \text{ km s}^{-1}$. This galaxy has not had its black-hole mass measured by any other technique so far but, given that most galaxies with bulges seem to have a black hole^{2,5}, from the σ – M_{BH} relation¹⁰ we can estimate $M_{\text{BH}} \approx 2 \times 10^8 M_\odot$. The sphere of influence (SOI) of a supermassive black hole (SMBH) would thus be $r_{\text{SOI}} = GM_{\text{BH}}/\sigma^2 \approx 20 \text{ pc}$ or $0.25''$ at the 16.4 Mpc distance of NGC 4526 (derived from surface-brightness fluctuations¹¹). NGC 4526 also has a large molecular gas reservoir (total molecular hydrogen mass $M_{\text{H}_2} = 3.69 \times 10^8 M_\odot$ (ref. 12)), previously mapped at low spatial resolution¹³ ($\sim 4''$). The molecular gas co-rotates with the stars¹⁴, and Hubble Space Telescope images show that it is coincident with regular dust lanes extending up to the galaxy centre, strongly suggesting that molecular gas with regular kinematics exists around the SMBH (ref. 7).

NGC 4526 was observed in the CO(2–1) line (230 GHz) using the Combined Array for Research in Millimetre Astronomy (CARMA) in A, B and C configurations¹⁵. The data were reduced in the standard manner^{16,17}, and more details are presented in Fig. 1. The spatial resolution achieved along the kinematic axis of the galaxy is $0.25''$ (20 pc), equal to the predicted SMBH SOI. The gas in this source seems to be distributed in a central component, and in an inner ring with spiral spurs that lead outwards to another ring at larger radii. A detailed analysis of the molecular gas morphology will be conducted in the future. We note, however, that the gas structures lie in the plane of the galaxy and seem to be regularly rotating; they are therefore likely to be dynamically cold and a good tracer of the potential¹⁸.

We created a grid of simulations of NGC 4526 spanning a range of SMBH masses (no SMBH, $M_{\text{BH}} = 3 \times 10^6$, and then from

$5 \times 10^7 M_\odot$ to $1.45 \times 10^9 M_\odot$ in linear steps) and I -band mass-to-light ratios ($M/L_I = 0.55$ – $6.15 M_\odot/L_\odot$ in linear steps). For full details of these simulations see section 1.1 in Supplementary Information. We fix the inclination of the gas disk¹⁹ ($i = 79^\circ$) and use an axisymmetric mass model of NGC 4526 (ref. 9) (carefully fitted to avoid contamination due to dust; see section 1.1.2 in Supplementary Information) to derive the circular-velocity curve expected from the luminous matter alone. The presence of a SMBH in NGC 4526 manifests itself as an inner Keplerian rise of the rotation curve (above that expected from luminous matter only). On larger angular scales such fast-rising rotation curves have been observed and have been used to infer the masses of central star clusters and bulges²⁰. We fitted these models to our observed data to determine whether such an excess due to a central dark mass is detectable in NGC 4526.

Figure 1a–c shows three different simulated position–velocity diagrams (PVDs) overlaid on the observed PVD of NGC 4526, with $\pm 1.15''$ insets. Figure 1a shows the best model with no SMBH, with a clear excess of high-velocity molecular gas at the centre. Figure 1b shows our overall best-fit model, clearly better reproducing the observed PVD at all radii, with $M_{\text{BH}} = 4.5 \times 10^8 M_\odot$ and $M/L_I = 2.65 M_\odot/L_\odot$. Figure 1c shows a model with a much larger SMBH, clearly incompatible with our data. Figure 1d–f shows the trace extracted from each observed PVD (as described in Supplementary Information), and the associated errors. The grey line shows the trace extracted from the models in the same way (with points denoting the values at the same radius as the observed points). The residuals between the data and the model at each position ($\Delta V = \text{data minus model}$) are shown in Fig. 1g–i. Clearly the best-fit model produces a significantly better fit to the data.

Figure 2 shows the χ^2 contours of our fits (conducted as described in Supplementary Information) as a function of M_{BH} and M/L_I . A clear global minimum is present at the best-fit values, with a minimum reduced χ^2 of 1.27. We define uncertainties for each of our fitted parameters from likelihood functions of each parameter (marginalized over the other) by finding the region containing, respectively, 68% and 99% of the probability. Our final best-fit values are $M_{\text{BH}} = 4.5^{+1.5}_{-1.3} \times 10^8 M_\odot$ with $M/L_I = 2.65 \pm 0.21 M_\odot/L_\odot$ (68% confidence level), and $4.5^{+4.2}_{-3.1} \times 10^8 M_\odot$ with $M/L_I = 2.65^{+0.56}_{-0.52} M_\odot/L_\odot$ (99% confidence level). These values are fully consistent within the 3σ uncertainties with the reported σ_e – M_{BH} relation¹⁰. Our formal uncertainties (0.14 dex and 0.5 dex at 1σ and 3σ , respectively) are also similar (and in many cases smaller) than the average uncertainties reported by other authors. For instance the mean 1σ error reported when using dynamical stellar and ionized gas techniques is ~ 0.6 dex (ref. 10) (however, the rather different systematic errors involved make direct comparison difficult).

The χ^2 contours show the usual degeneracy between SMBH mass and mass-to-light ratio found in other studies. To fit the data without a SMBH, a negative M/L gradient would be necessary, with older stellar populations dominating in the inner parts (high M/L) and young stars

¹European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching-bei-München, Germany. ²Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK. ³Centre for Astrophysics Research, University of Hertfordshire, Hatfield, Herts AL1 9AB, UK. ⁴Department of Astronomy, University of California, Berkeley, California 94720, USA.

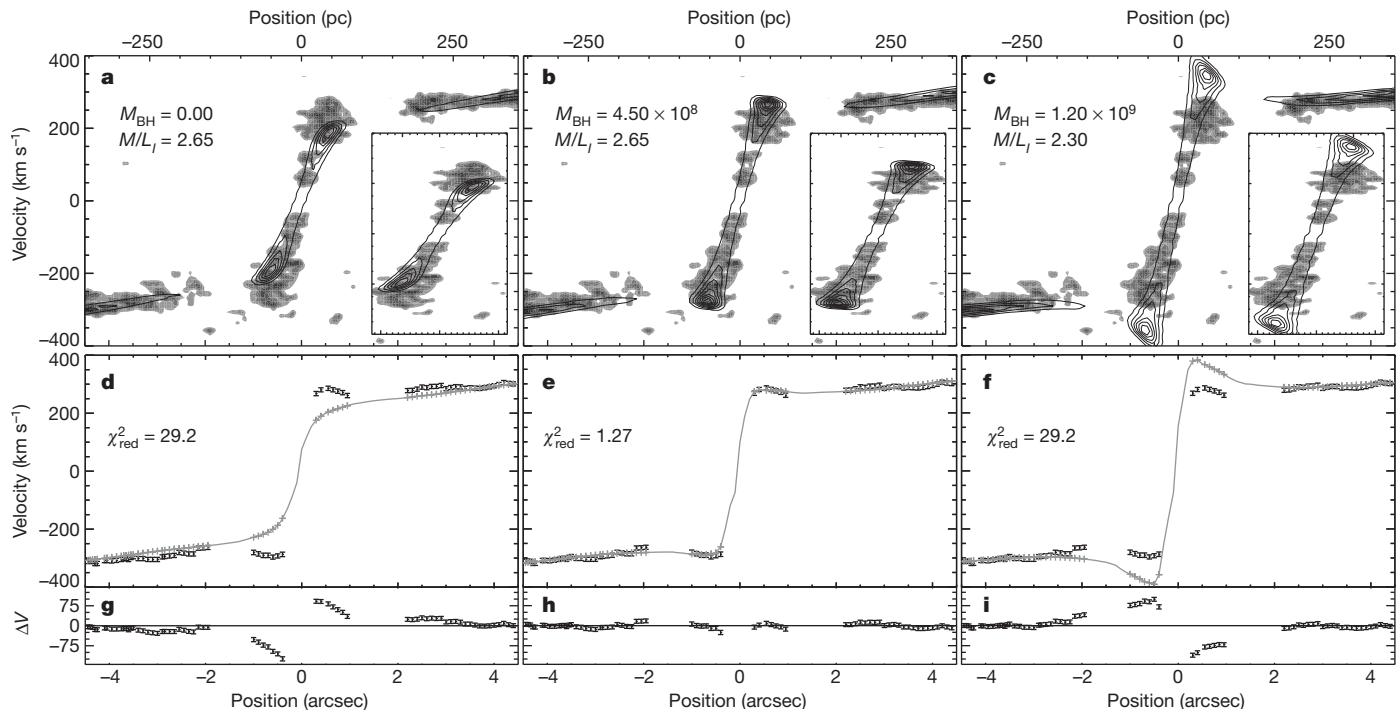


Figure 1 | NGC 4526 kinematic models and data. **a–c**, Model PVDs (black contours), overlaid on the observed CO(2–1) PVD (greyscale). This PVD was created from our CO(2–1) observations of NGC 4526 from CARMA. The synthesized beam size achieved in these observations was $0.27'' \times 0.17''$, and the velocity channel width was 10 km s^{-1} . The final fully reduced and calibrated data cube had a root-mean-square noise of 2.88 mJy per beam. The PVD was created by rotating the data cube to align the kinematic major axis of the molecular gas with the x axis, and then summing over one beam width around the axis in the y direction. The spatial resolution achieved in the PVD was $0.25''$ (20 pc), equal to the predicted SMBH SOI. Our results do not depend on the method used to extract the PVD. Shown are the best model with no SMBH

(**a**), the overall best-fit model (**b**) and a model with an overweight SMBH (**c**). The model M_{BH} and M/L_I are indicated in the top left corner of each panel, and an inset of the central $\pm 1.15''$ is shown in the bottom right corner.

d–f, Black points show the trace extracted from the respective observed PVD, and associated standard errors. The grey line shows the trace extracted from the models, and the grey crosses denote the value of the trace at the same radius as the observed points. **g–i**, Residuals between the model and data at each position ($\Delta V = \text{data minus model; km s}^{-1}$). The error bars in **d–f** and **g–i** correspond to the formal uncertainties in fitting the trace (see section 1.2 in Supplementary Information), added in quadrature with two factors of 5 km s^{-1} (to account for the finite velocity channel width in both the data and model).

at large radii (low M/L). Although the real M/L is unlikely to be exactly constant, this is opposite to the trend reported by reddening free-stellar-population studies²¹ and that expected from the presence of molecular gas and star formation in the central regions²².

The ability to determine a black-hole mass accurately by using molecular gas can be affected by many of the same issues that affect

measurements of ionized gas. Turbulent motions are in general small in molecular gas, but they could conceivably increase around a black hole. Similarly, if the inner gas were to be misaligned from the stellar body, our mass estimate would be systematically affected in a way that is degenerate with a change in M/L . In this galaxy, however, we find no evidence that the velocity dispersion increases in the inner regions (as described in section 1.1.1 in Supplementary Information), and constrain the inner gas to be aligned with the stellar body within $< 3^\circ$. Such a misalignment could change velocities by an insignificant amount ($\sim 3 \text{ km s}^{-1}$; see section 1.1.2 in Supplementary Information). The presence of dust could also cause mass models to underestimate the contribution of luminous matter to the potential. We again believe this should not introduce significant errors in this object, because of our careful treatment of dust in the mass model (see section 1.1.2 in Supplementary Information). Future studies with this technique should choose their targets to minimize the impact of such effects, but they can also include warps and turbulent motions in their gas disk models²³ and use near-infrared photometry as done in previous studies of dusty objects²³.

The use of molecular gas as a kinematic tracer holds great promise to increase the total number of SMBH mass measurements in galaxies of all types where CO is detected. Angular resolutions of up to $0.15''$ can be achieved with current millimetre-wavelength interferometers, allowing one, for instance, to resolve the SOI of a $2 \times 10^8 M_\odot$ SMBH in a galaxy with $\sigma_e = 200 \text{ km s}^{-1}$ only out to about 30 Mpc . The next generation of millimetre-wavelength interferometers will have an angular resolution that is more than an order of magnitude greater (for example, $\sim 6 \text{ mas}$ at 690 GHz with ALMA), as well as greatly increased sensitivity. In less than 5 h of integration time with ALMA

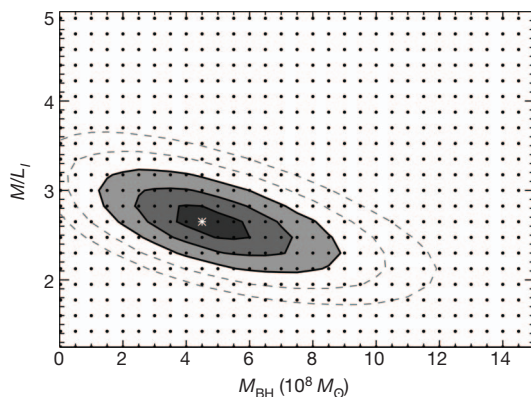


Figure 2 | NGC 4526 SMBH mass uncertainties. $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\text{minimum}}$ contours of our fits to the CO(2–1) PVD, as a function of the two free parameters M_{BH} and M/L_I . The model grid is shown with black dots and the overall best-fit model by a white star. The solid shaded contours with black lines correspond to the 1σ – 3σ levels with one degree of freedom ($\Delta\chi^2 = 1, 4, 9$) with good models in the darkest areas. The 4σ and 5σ levels ($\Delta\chi^2 = 16, 25$) are shown by dashed grey lines.

one could achieve the same sensitivity and linear resolution as the observations presented here, in a galaxy 75 Mpc away. Galaxies of lower mass (with smaller SMBHs) will also become accessible. For instance, the SOI of a Milky-Way-like SMBH²⁴ ($M_{\text{BH}} = 4 \times 10^6 M_{\odot}$, $\sigma_e = 105 \text{ km s}^{-1}$) will be resolved up to ~ 50 Mpc.

Measurements using a single technique will thus be possible over the entire range of the σ_e – M_{BH} relation, leading to much-needed reduced uncertainties in the slope and normalization of black-hole–galaxy relations. This will allow the study of a large number of spiral galaxies that cannot currently be probed easily with stellar dynamical techniques. Furthermore, it will also provide access to a larger number of systems than ionized-gas techniques (currently limited by the $\sim 0.05''$ resolution of the Hubble Space Telescope). Even considering that only some of the accessible objects will have suitably relaxed and centrally peaked distributions of molecular gas, measuring SMBH masses using molecular gas will be possible in many hundreds of spiral (and early-type) galaxies, many times more than are possible with conventional stellar and ionized-gas tracers today. This technique could also be extended to the use of other spectral lines, such as the very bright atomic gas cooling lines at higher frequencies (to increase the spatial resolution even further) and optically thin transitions such as ^{13}CO (which would allow determinations of black-hole masses even in systems that are exactly edge on).

Received 11 September; accepted 27 November 2012.

Published online 30 January 2013.

- Magorrian, J. *et al.* The demography of massive dark objects in galaxy centers. *Astron. J.* **115**, 2285–2305 (1998).
- Graham, A., Erwin, P., Caon, N. & Trujillo, I. A correlation between galaxy light concentration and supermassive black hole mass. *Astrophys. J.* **563**, L11–L14 (2001).
- Silk, J. & Rees, M. Quasars and galaxy formation. *Astron. Astrophys.* **331**, L1–L4 (1998).
- Ferrarese, L. & Ford, H. Supermassive black holes in galactic nuclei: past, present and future research. *Space Sci. Rev.* **116**, 523–624 (2005).
- Sarzi, M. *et al.* Supermassive black holes in bulges. *Astrophys. J.* **550**, 65–74 (2001).
- Barth, A. J. *et al.* Evidence for a supermassive black hole in the S0 galaxy NGC 3245. *Astrophys. J.* **555**, 685–708 (2001).
- Ho, L. C. *et al.* An efficient strategy to select targets for gas dynamical measurements of black hole masses using the Hubble Space Telescope. *Publ. Astron. Soc. Pacif.* **114**, 137–143 (2002).
- Lo, K. Y. Mega-masers and galaxies. *Annu. Rev. Astron. Astrophys.* **43**, 625–676 (2005).
- Cappellari, M. *et al.* The SAURON project. IV. The mass-to-light ratio, the virial mass estimator and the fundamental plane of elliptical and lenticular galaxies. *Mon. Not. R. Astron. Soc.* **366**, 1126–1150 (2006).
- Gultekin, K. *et al.* Determination of the intrinsic scatter in the M – σ and M – L relations. *Astrophys. J.* **698**, 198–221 (2009).
- Tonry, J. *et al.* The SBF survey of galaxy distances. IV. SBF magnitudes, colors, and distances. *Astrophys. J.* **546**, 681–693 (2001).
- Combes, F., Young, L. M. & Bureau, M. Molecular gas and star formation in the SAURON early-type galaxies. *Mon. Not. R. Astron. Soc.* **377**, 1795–1807 (2007).
- Young, L. M., Bureau, M. & Cappellari, M. Structure and kinematics of molecular disks in fast-rotator early-type galaxies. *Astrophys. J.* **676**, 317–334 (2008).
- Davis, T. A. *et al.* The ATLAS^{3D} project. X. On the origin of the molecular and ionized gas in early-type galaxies. *Mon. Not. R. Astron. Soc.* **417**, 882–899 (2011).
- Bock, D. *et al.* CARMA: Combined Array for Research in Millimeter-Wave Astronomy. *Proc. SPIE* **6267**, 13 (2006).
- Alatalo, K. *et al.* The ATLAS^{3D} project. XVIII. CARMA CO imaging survey of early-type galaxies. Preprint at <http://arxiv.org/abs/1210.5524v1> (2012).
- Sault, R., Teuben, P. & Wright, M. A retrospective view of MIRIAD. *Astron. Soc. Pacif. Conf. Ser.* **77**, 433–436 (1995).
- Davis, T. A. *et al.* The ATLAS^{3D} Project. XIV. The extent and kinematics of molecular gas in early-type galaxies. Preprint at <http://arxiv.org/abs/1211.1011v1> (2012).
- Davis, T. A. *et al.* The ATLAS^{3D} project. V. The CO Tully–Fisher relation of early-type galaxies. *Mon. Not. R. Astron. Soc.* **414**, 968–984 (2011).
- Sofue, Y., Koda, J., Nakanishi, H. & Onodera, S. The Virgo high-resolution CO survey. II. Rotation curves and dynamical mass distributions. *Publ. Astron. Soc. Jpn* **55**, 59–74 (2003).
- Kuntschner, H. *et al.* The SAURON project. XVII. Stellar population analysis of the absorption line strength maps of 48 early-type galaxies. *Mon. Not. R. Astron. Soc.* **408**, 97–132 (2010).
- Crocker, A. F., Bureau, M., Young, L. M. & Combes, F. Molecular gas and star formation in early-type galaxies. *Mon. Not. R. Astron. Soc.* **410**, 1197–1222 (2011).
- Neumayer, N. *et al.* The central parsecs of Centaurus A: high-excitation gas, a molecular disk, and the mass of the black hole. *Astrophys. J.* **671**, 1329–1344 (2007).
- Ghez, A. *et al.* Measuring distance and properties of the Milky Way's central supermassive black hole with stellar orbits. *Astrophys. J.* **689**, 1044–1062 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements The research leading to these results has received funding from the European Community's Seventh Framework Programme. M.B. is supported by the rolling grants 'Astrophysics at Oxford' and from the UK Research Councils. M.C. acknowledges support from a Royal Society University Research Fellowship. M.S. acknowledges support from a Science and Technology Facilities Council Advanced Fellowship. Support for the construction of CARMA was derived from the states of California, Illinois and Maryland, the James S. McDonnell Foundation, the Gordon and Betty Moore Foundation, the Kenneth T. and Eileen L. Norris Foundation, the University of Chicago, the Associates of the California Institute of Technology, and the National Science Foundation. Ongoing development and operations of CARMA are supported by the National Science Foundation under a cooperative agreement, and by the CARMA partner universities.

Author Contributions T.A.D. prepared and reduced the observations, and created the modelling tool. T.A.D. and M.B. prepared the manuscript. M.C. created the mass model. All authors discussed the results and implications and commented on the manuscript at all stages.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A.D. (tdavis@eso.org).

Generation of electron Airy beams

Noa Voloch-Bloch¹, Yossi Lereah¹, Yigal Lilach¹, Avraham Gover¹ & Ady Arie¹

Within the framework of quantum mechanics, a unique particle wave packet exists¹ in the form of the Airy function^{2,3}. Its counter-intuitive properties are revealed as it propagates in time or space: the quantum probability wave packet preserves its shape despite dispersion or diffraction and propagates along a parabolic caustic trajectory, even though no force is applied. This does not contradict Newton's laws of motion, because the wave packet centroid propagates along a straight line. Nearly 30 years later, this wave packet, known as an accelerating Airy beam, was realized⁴ in the optical domain; later it was generalized to an orthogonal and complete family of beams⁵ that propagate along parabolic trajectories, as well as to beams that propagate along arbitrary convex trajectories⁶. Here we report the experimental generation and observation of the Airy beams of free electrons. These electron Airy beams were generated by diffraction of electrons through a nano-scale hologram^{7–9}, which imprinted on the electrons' wavefunction a cubic phase modulation in the transverse plane. The highest-intensity lobes of the generated beams indeed followed parabolic trajectories. We directly observed a non-spreading electron wavefunction that self-heals¹⁰, restoring its original shape after passing an obstacle. This holographic generation of electron Airy beams opens up new avenues for steering electronic wave packets like their photonic counterparts, because the wave packets can be imprinted with arbitrary shapes⁵ or trajectories⁶.

Curved light is an intriguing caustic phenomenon² common in nature and in everyday life; examples range from rainbows³ to the bright light patterns that appear on the sea floor when sun shines on rippling water waves². Caustics, already studied in the nineteenth century, led George Biddell Airy to discover the Airy function³. When the quantum mechanical probability density of a particle is imposed with the initial shape of an Airy function, the regions of maximum probability, that is, the intensity peaks of the Airy function, preserve their shape¹¹ and stay localized around parabolic trajectories in space, similar to that of a freely propagating projectile experiencing a transverse accelerating force. As already pointed out, this accelerating Airy wave packet¹ does not contradict Ehrenfest's theorem (embodying Newton's second law of motion), as the wave packet centroid travels along a straight path. The evolution of a light beam in space corresponds to the paraxial Helmholtz equation, which resembles the Schrödinger equation. This analogy led to the discovery and the experimental realization of the optical Airy beams⁴. The intriguing propagation dynamics of Airy beams is a caustic wave phenomenon that can be understood by ray analysis: the wave packet is composed of a family of rays that coalesce along a curved boundary¹². The local angular momentum of Airy beams changes as they propagate, but the total momentum and energy are conserved¹³. Various applications followed the discovery of optical Airy beams, including microparticle manipulation¹⁴, generation of plasma channels in air and water¹⁵, surface Airy plasmons^{16–18} and applications in lasers^{19,20} and in nonlinear optics²¹. All of these rely on the Airy form of the wave packet of photons.

The spatial evolution of the envelope, Ψ , of an electron's wavefunction can be expressed using the paraxial Helmholtz equation (Supplementary Information, section 1):

$$\left(\nabla_{\perp}^2 + 2ik_B \frac{\partial}{\partial z}\right)\Psi = 0$$

where $\nabla_{\perp}^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the transverse derivative and $k_B = p/\hbar = 2\pi/\lambda_B$ is the de Broglie wavenumber of the electron (\hbar , Planck's constant divided by 2π). This equation has the same form as that of the Schrödinger equation. However, rather than describing the time evolution of the electron's wavefunction, it describes the wavefunction's evolution as it propagates in space. This is another manifestation of the analogy, widely used in optics, between beam diffraction in space and pulse dispersion in time. When the initial wavefunction of the electron is an Airy function (Ai), $\Psi(x, y, z=0) = \text{Ai}(x/x_0)\text{Ai}(y/y_0)$, where x_0 and y_0 are characteristic length scales, the general solution for the wave packet is^{1,10,14}

$$\begin{aligned} \Psi(x, y, z) = & \text{Ai}\left(\frac{x}{x_0} - \frac{z^2}{4k_B^2 x_0^4}\right) \text{Ai}\left(\frac{y}{y_0} - \frac{z^2}{4k_B^2 y_0^4}\right) \\ & \times \exp\left(i\frac{xz}{2k_B x_0^3} - i\frac{z^3}{12k_B^3 x_0^6}\right) \\ & \times \exp\left(i\frac{yz}{2k_B y_0^3} - i\frac{z^3}{12k_B^3 y_0^6}\right) \end{aligned} \quad (1)$$

It is then clear from equation (1) that $|\Psi|^2$ preserves its shape and follows a parabolic trajectory. The parabolic trajectory is described by²² $x(z) = z^2/4k_B^2 x_0^3$, $y(z) = z^2/4k_B^2 y_0^3$. The ideal Airy beam carries an infinite amount of energy, whereas in practice the Airy beam is truncated, having a finite energy. The finite Airy beam is obtained by multiplying the Airy function with an exponential or Gaussian window⁴. Nevertheless, over a finite distance, the finite Airy beam has all the special characteristics of the infinite Airy beam such as slow diffraction, curved trajectory and self-healing.

In optics, finite Airy beams were experimentally obtained by passing a Gaussian beam through a phase mask imprinting a cubic phase modulation in the transverse direction⁴, and then doing an optical Fourier transform (the Fourier transform of a function having a cubic phase modulation results in an Airy function). Here we use electrons instead of photons. These electrons pass through a hologram that adds a transverse cubic phase, $\exp(i\phi(x, y)) = \exp(ic_x x^3)\exp(ic_y y^3)$, to the wavefunction. Our hologram design method was to construct a binary diffraction grating with the following shape^{23,24}:

$$h(x, y) = \frac{1}{2} h_0 (\text{sgn}\{\cos[2\pi x/A + c_x x^3 + c_y y^3] + D_{\text{cycle}}\} + 1)$$

In this way, a cubic phase is imposed on a carrier frequency. The carrier period is A , h_0 is the ridge height of the binary phase mask and D_{cycle} is an arbitrary duty cycle factor. When electrons (or light) diffract from the binary structure, it decomposes into different diffraction orders; the complex amplitude of the m th-order diffracted beam is proportional to $\exp(im\phi(x, y))$. In the special case of a cubic phase modulation, each diffraction order is superimposed with a different cubic phase and so has an amplitude proportional to $\exp(im(c_x x^3 + c_y y^3))$. Each order therefore propagates along a different parabolic trajectory with a quadratic coefficient proportional to $1/m$ (Supplementary Information,

¹Department of Physical Electronics, School of Electrical Engineering, Fleischman Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel.

section 2). We name this diffraction pattern the ‘Airy lattice’ (Figs 1 and 2). The general parametric equations of the trajectories of order m are

$$\begin{aligned}x_m(z) &= -\frac{1}{m} \frac{1}{4k_B^2 x_0^3} z^2 + mz \frac{G}{k_B} + mf \frac{G}{k_B} \\y_m(z) &= \frac{1}{m} \frac{1}{4k_B^2 y_0^3} z^2\end{aligned}\quad (2)$$

where $G = 2\pi/\Lambda$ is the reciprocal lattice vector of the transmission hologram and f is the focal length of the lens. Note that the y trajectories include only a quadratic term, whereas the x trajectories contain both a quadratic term and a linear term, owing to the modulation of the carrier frequency. The Airy lattice is a novel type of lattice that diffracts but does not spread. Although each order diffracts in a different direction, each order stays localized and anomalously bends.

In our experiment, we use a field-emission-gun transmission electron microscope (FEG-TEM), operating at 200 keV. The de Broglie wavelength, including relativistic correction in this case, is approximately 2.5 pm. Therefore, to generate the inner structure of Airy beams as well as the spatial separation of different orders, we used nanoscale holograms (Supplementary Information, section 3). The micrographs of these nanoscale holograms are shown in Fig. 3b, c. The Fourier transform of the modulated wavefunction is done using a set of

magnetic lenses, so that an electron Airy beam is obtained in the back focal plane of the FEG-TEM. This method is analogous to the method used to generate optical Airy beams⁴, the only difference being that we manipulate electrons rather than photons. Some of the results from the Airy profiles are shown in Fig. 3d and Fig. 3e. Because the measurement plane is located at a fixed position in our FEG-TEM, we varied the focal lengths of the magnetic lenses to observe the formation and evolution of the electron Airy wave packet in space (Fig. 3a). The experiment details are given in Supplementary Information, section 4.

In the first experiment, we explored the parabolic trajectories of electron Airy beams (Fig. 1). We recorded profiles using an electron beam with a relatively large area ($100\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$). Figure 1a shows the numerical (simulated using the beam propagation numerical method²¹) and the experimental transverse profiles of the diffracted electrons at several propagation distances behind the Fourier plane. The viewed diffracting orders are +1, -1, +2 and -2. As seen, Airy orders +1 and -1 anomalously converge in x and eventually collide, unlike the normal diffraction of Bragg peaks, which linearly diverge. As expected, the orders diverge in y . Unlike orders +1 and -1, orders +2 and -2 are almost constant in x and only slightly diverge in y . We traced the trajectories of the highest intensity lobe, of orders +1 and +2. The projections of the deflections of the orders into the x - z and y - z planes are shown in Fig. 1b. The values of $x(z)$ and $y(z)$ are given by pixel numbers, multiplied by $25\text{ }\mu\text{m}$ per pixel. The measured parabolic

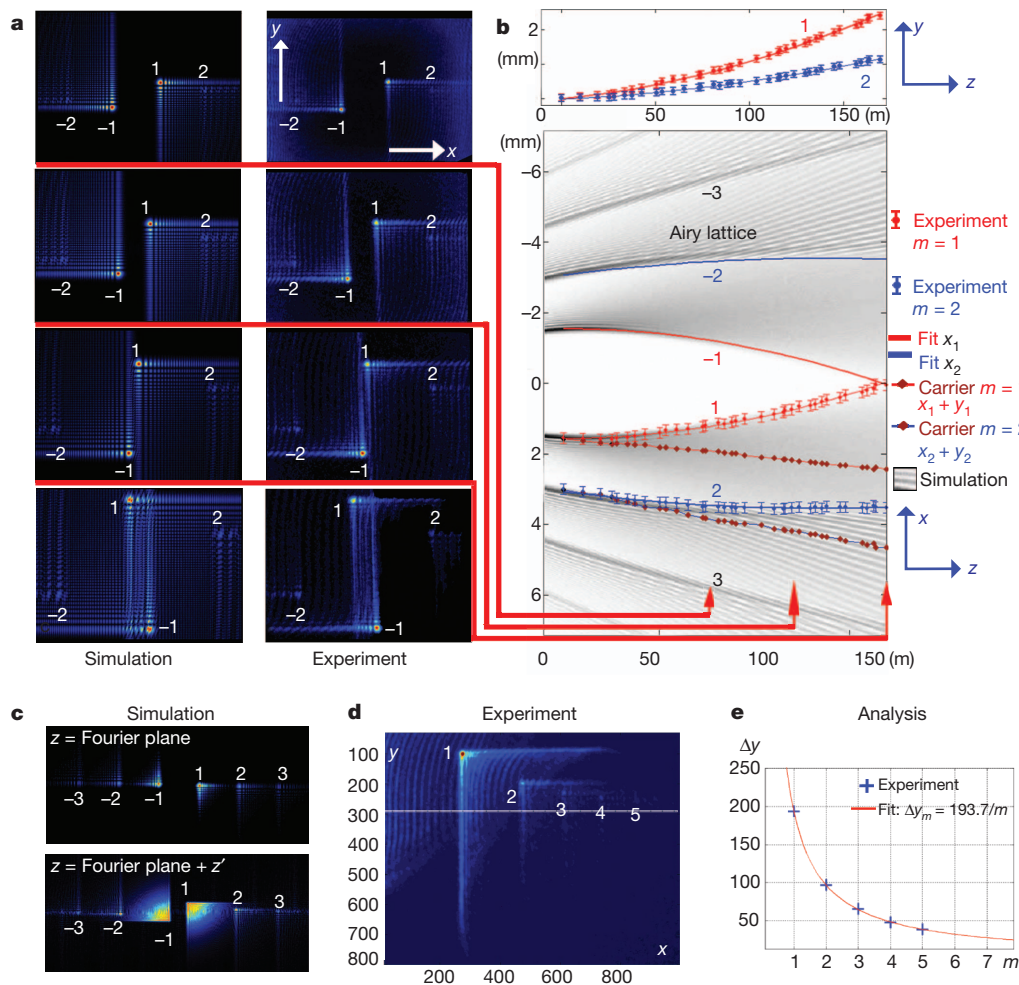


Figure 1 | Exploring the propagation dynamics of the Airy lattice.

a, Comparison of micrographs of the transverse (x, y) profiles of the Airy lattice with simulation results, revealing the propagation dynamics of orders +1, -1, +2 and -2. **b**, The simulated trajectories of multiple Airy orders projected into the x - z plane versus the experimental results for orders +1 and +2 (y - z plane

also shown). The error bars indicate the range of the estimated systematic error in the measurements. **c**, Propagation of the Airy lattice. **d**, Experimental micrograph revealing the first five Airy orders. **e**, The deflection in y of each order m is proportional to $1/m$.

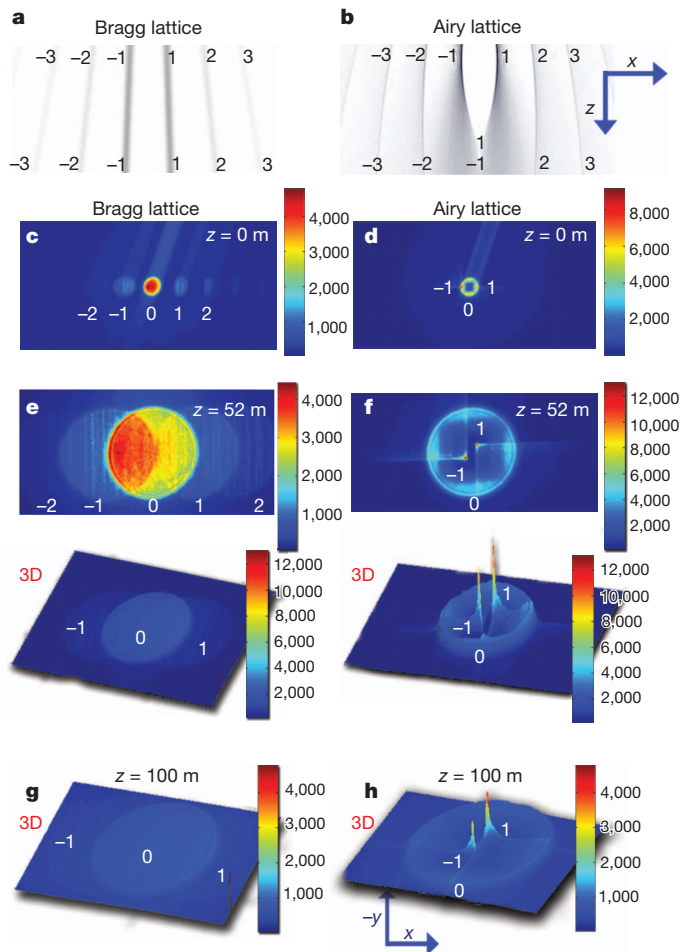


Figure 2 | Comparison between electrons diffracting from an Airy grating and the electrons diffracting from a reference periodic Bragg grating. **a**, The diffraction from a Bragg lattice is normally outward at an angle $\alpha_m = m\lambda_B/\Lambda$. **b**, The diffraction from an Airy grating is anomalous because the lattice peaks are curved inward. **c–h**, Experimental profile micrographs of different propagation planes. Notice that the zeroth order looks the same in both the Airy lattice and the Bragg lattice. The Airy orders are very localized and maintain high intensities compared with the Bragg orders.

trajectories are in the shape given by the analytical trajectories in equation (2). We note that the summation of the deflection of the carrier and the deflection in x is equal to the deflection in y . Also, the transverse linear coefficient increases by a factor of m , whereas the quadratic term decreases by a factor of $1/m$. This is why Airy beams of high order have almost linear trajectories like the Bragg peaks. This also explains the experimental results and the differences in propagation between orders 1 and -1 (which eventually collide in x) and orders 2 and -2 . The fitting coefficients for the experimental results are presented in Table 1.

Another way to find the quadratic coefficient of the parabolic trajectory is by using only a single-profile picture at a fixed distance, z , from the Fourier plane (Fig. 1c–e). A single-profile picture reveals several Airy orders. Each Airy order diffracts with a different parabolic trajectory, having a different quadratic coefficient which is proportional to $1/m$. The deflection in y of each order in a single propagation plane (z has a fixed value) is $\Delta y = C_1/m$, where C_1 is a quadratic coefficient. This causes the Airy lattice to change its proportions. The diffraction orders that were located equidistantly in the Fourier plane, $z = 0$, diverge non-uniformly, and their proportions are not maintained (Fig. 1c and Supplementary Information, section 2). This is another demonstration of the parabolic propagation dynamics, which have been verified experimentally (Fig. 1d). The fitting curve

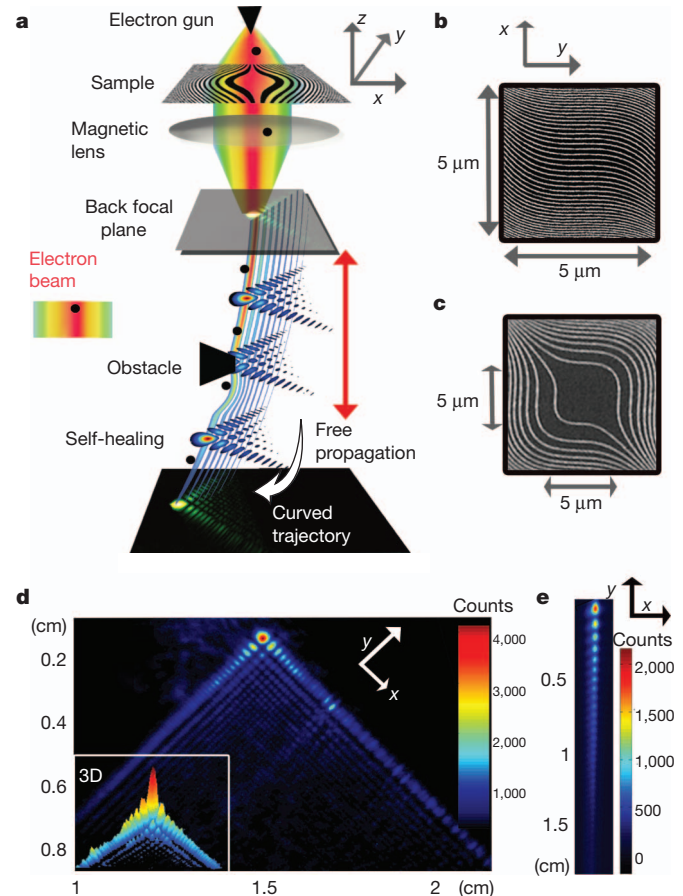


Figure 3 | Holographic generation of an electron Airy beam. **a**, An electron beam is transmitted through a nanofabricated hologram, with a cubic phase modulation. It is then focused by a magnetic lens. The Airy wave packet is formed at the back focal plane and recorded as it propagates. The electron Airy beam is a shape-preserving beam that evolves along a curved trajectory. It also self-heals after encountering obstacles. **b**, **c**, TEM micrographs of the nanoscale Airy beam: two-dimensional Airy on a spatial carrier frequency (**b**); two-dimensional Airy without a carrier (**c**). **d**, **e**, Experimental wave packet micrographs of two-dimensional (**d**) and one-dimensional (**e**) electron Airy beams.

of the experimental results presented in Fig. 1e is $\Delta y = 193.7/m$, matching the theoretical prediction. Detailed explanation of the parabolic trajectories measurement is given in Supplementary Information, sections 2 and 5. The collision between Airy beams represents a new way of interfering electron beams. The electron beam is separated into two diffraction orders, but these beams re-merge owing to their opposite directions (Fig. 1a, b). The measured deflection in x and y was 2.5 mm over an effective distance of 150 m.

We compared the electrons diffracting from an Airy grating with the electrons diffracting from a reference periodic Bragg grating. We simulated the evolution of electrons (Fig. 2a) diffracted from a periodic Bragg grating and from an Airy grating with the same carrier period. (Only in this case did we use different input beam areas for Airy and Bragg lattices, to visualize the trajectories.) As expected, the diffraction from a Bragg lattice is linear and outward; however, the diffraction

Table 1 | Quadratic fit results

| Coefficient | C_1 (m^{-1}) | C_2 | C_3 (m) |
|-------------|---|---|--|
| $m = 1$ | | | |
| Δx | $-1.022\text{e}^{-7} \pm 1.51\text{e}^{-9}$ | $7.141\text{e}^{-6} \pm 2.7\text{e}^{-7}$ | $1.43\text{e}^{-3} \pm 4\text{e}^{-5}$ |
| Δy | $1.022\text{e}^{-7} \pm 1.51\text{e}^{-9}$ | 0 | $-1.41\text{e}^{-4} \pm 2\text{e}^{-5}$ |
| $m = 2$ | | | |
| Δx | $-4.869\text{e}^{-8} \pm 1.02\text{e}^{-9}$ | $1.312\text{e}^{-5} \pm 2.7\text{e}^{-7}$ | $2.68\text{e}^{-3} \pm 3.75\text{e}^{-5}$ |
| Δy | $4.869\text{e}^{-8} \pm 1.02\text{e}^{-9}$ | 0 | $-8.59\text{e}^{-5} \pm 1.42\text{e}^{-5}$ |

Fits to $\Delta x, \Delta y = C_1 z^2 + C_2 z + C_3$ and 95% confidence intervals calculated from trough curve fitting.

Table 2 | Comparison between non-spreading Airy beams and spreading Bragg peaks

| Propagation distance, z (m) | FWHM Bragg (μm) | FWHM Airy (μm) | FWHM Bragg/Airy |
|-------------------------------|------------------------------|-----------------------------|-----------------|
| 0 | 1,125 | 102 | 11 |
| 52 | 5,785 | 104 | 58 |
| 100 | 8,125 | 110 | 81 |

FWHMs of the diffracted beams.

from the Airy lattice is anomalously bent (inwards for low-index diffraction orders). In the second experiment, we recorded profiles (Fig. 2) using a relatively narrow input beam area ($10\ \mu\text{m} \times 10\ \mu\text{m}$), letting the zero-order diffraction peak spread out to enable visualization of the difference between the evolution of spreading Bragg peaks and that of the shape-preserving Airy lattice peaks. The zeroth order of the Airy lattice is the only one that is not imprinted with a cubic phase; thus, it spreads in the normal manner. The zeroth order appears as a large circular spot in the middle of the frame. We can also see diffraction orders $+1$, -1 , $+2$ and -2 . At an effective distance of 100 m, the diffraction patterns from the Bragg lattice (as well as from the zeroth order of the Airy lattice) spread out linearly, become very large and decay in intensity. However, for $m \neq 0$ the curved Airy peaks stay confined to a very small area and maintain their high intensity. This difference is emphasized when calculating the full-widths at half-maximum (FWHMs) of the diffracted orders (Table 2).

We also measured the self-healing properties of electron Airy beams (Fig. 4). For this purpose, we used a glass wire placed in the diffraction plane. This wire is conventionally used as a bi-prism in electron holography²⁵, but in our case it was used simply to block parts of the beam. Increasing the current of the objective lens raised the Airy beam above the wire. Then we adjusted the wire to block Airy orders -1 and $+1$ simultaneously. We blocked the different orders in a slightly different manner. We then gradually increased the current of the diffraction lens and observed self-healing of the two orders¹⁰. The wave packets reconstructed their shapes after passing the blocking wire. As shown, the two orders self-healed differently. Order 1 self-healed faster than order -1 . We also simulated the self-healing process and the numerical results are in a good agreement with the experimental results.

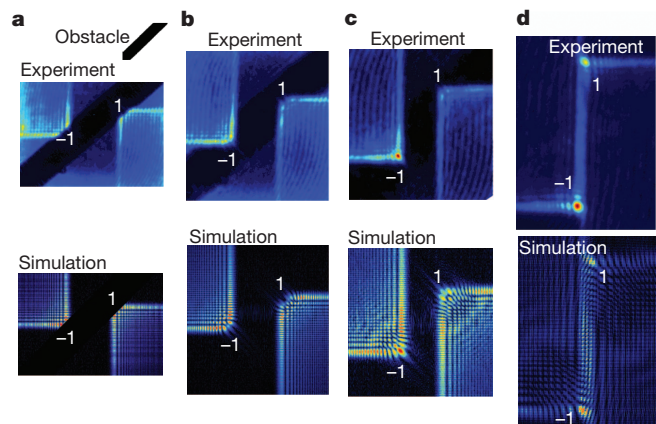
We have experimentally observed a non-spreading electron Airy wave packet whose highest-intensity quantum probability peaks bend in the absence of an external force. Our technique for generating electron Airy beams is analogous to the optical method⁴, but makes use of recent advances in electron beam shaping⁷ and nanoscale hologram fabrication techniques^{8,9}. We measured the wave packet trajectories in two different ways. Our method of holographically generating

electron Airy beams suggests a general means of manipulating particles' wave packet trajectories by engineering their initial probability density wave functions. Our experimental results include demonstration of electron wave packet trajectories that show shape and size preservation over an effective length of more than 100 m. Such non-spreading Airy electron wave packets may be useful in improving the resolution properties of TEM imaging, because they have an extremely large depth of focus. Furthermore, we have shown that these electron wave packets self-heal and reconstruct their original shape after passing an obstacle. An interesting feature of these beams is that different diffraction orders with quadratic terms of opposite sign can be merged and possibly used as a new type of electron wave interferometer. The holographic generation of Airy electron beams can be generalized because recent studies have shown that optical Airy beams can be designed with arbitrary spatial shapes⁵ and can propagate along arbitrary trajectories⁶. It may also be possible to use these beams to study electron spin-orbit interaction effects in the relativistic regime, similarly to recent studies of electron vortices²⁶. Finally, there may be other ways of manipulating and shaping the trajectories and the self-healing properties of Airy beams, based on the ability to externally influence electrons using magnetic or electric potentials²⁷.

METHODS SUMMARY

Our analytical method of designing nanoscale holograms for electrons, imposed with any phase, is described in Supplementary Information, section 2. Our technique for determining the propagation dynamics of an electron beam, which we showed can be modelled numerically, is reported in Supplementary Information, section 5. We also developed a way of imaging electron wave packets with high magnification (Supplementary Information, section 4) and a way of manufacturing holographic masks for electrons with high writing resolution (Supplementary Information, section 3).

Received 2 August; accepted 4 December 2012.

**Figure 4 | Self-healing properties of electron Airy wave packets.**

a–d, Experimental profile photograph versus numerical simulation at different propagation planes after Fourier plane. The diffracted orders 1 and -1 are blocked differently and their self-healing processes are different. The two orders are blocked with a glass wire.

- Berry, M. V. & Balazs, N. L. Nonspreading wave packets. *Am. J. Phys.* **47**, 264–267 (1979).
- Berry, M. & Upstill, C. *Catastrophe Optics: Morphologies of Caustics and their Diffraction Patterns* (Elsevier, 1980).
- Airy, G. B. On the intensity of light in the neighbourhood of a caustic. *Trans. Camb. Phil. Soc.* **6**, 379–403 (1838).
- Siviloglou, G. A., Broky, J., Dogariu, A. & Christodoulides, D. N. Observation of accelerating Airy beams. *Phys. Rev. Lett.* **99**, 213901 (2007).
- Bandres, M. A. Accelerating parabolic beams. *Opt. Lett.* **33**, 1678–1680 (2008).
- Greenfield, E., Segev, M., Walasik, W. & Raz, O. Accelerating light beams along arbitrary convex trajectories. *Phys. Rev. Lett.* **106**, 213902 (2011).
- Uchida, M. & Tonomura, A. Generation of electron beams carrying orbital angular momentum. *Nature* **464**, 737–739 (2010).
- Verbeeck, J., Tian, H. & Schattschneider, P. Production and application of electron vortex beams. *Nature* **467**, 301–304 (2010).
- McMorran, B. J. *et al.* Electron vortex beams with high quanta of orbital angular momentum. *Science* **331**, 192–195 (2011).
- Broky, J., Siviloglou, G. A., Ariste, D. & Christodoulides, D. N. Self-healing properties of optical Airy beams. *Opt. Express* **16**, 12880–12891 (2008).
- Durnin, J., Miceli, J. J. & Eberly, J. H. Diffraction-free beams. *Phys. Rev. Lett.* **58**, 1499–1501 (1987).
- Kaganovsky, Y. & Heyman, E. Wave analysis of Airy beams. *Opt. Express* **18**, 8440–8452 (2010).
- Sztul, H. I. & Alfano, R. R. The Poynting vector and angular momentum of Airy beams. *Opt. Express* **16**, 9411–9416 (2008).
- Baumgartl, J., Mazilu, M. & Dholakia, K. Optically mediated particle clearing using Airy wave packets. *Nature Photon.* **2**, 675–678 (2008).
- Polynkin, P., Kolesik, M., Moloney, J., Siviloglou, G. A. & Christodoulides, D. N. Curved plasma channel generation using ultraintense Airy beams. *Science* **324**, 229–232 (2009).
- Salandrino, A. & Christodoulides, D. N. Airy plasmon: a nondiffracting surface wave. *Opt. Lett.* **35**, 2082–2084 (2010).
- Minovich, A. *et al.* Generation and near-field imaging of Airy surface plasmons. *Phys. Rev. Lett.* **107**, 116802 (2011).
- Zhang, P. *et al.* Plasmonic Airy beams with dynamically controlled trajectories. *Opt. Lett.* **36**, 3191–3193 (2011).
- Porat, G., Dolev, I., Barlev, O., & Arie, A. Airy beam laser. *Opt. Lett.* **36**, 4119–4121 (2011).
- Longhi, S. Airy beams from a microchip laser. *Opt. Lett.* **36**, 716–718 (2011).
- Ellenbogen, T., Bloch, N. V., Ganany-Padowicz, A. & Arie, A. Nonlinear generation and manipulation of Airy beams. *Nature Photon.* **3**, 395–398 (2009).
- Siviloglou, G. A., Broky, J., Dogariu, A. & Christodoulides, D. N. Ballistic dynamics of Airy beams. *Opt. Lett.* **33**, 207–209 (2008).

23. Dai, H. T., Sun, X. W., Luo, D. & Liu, Y. J. Airy beams generated by a binary phase element made of polymer-dispersed liquid crystals. *Opt. Express* **17**, 19365–19370 (2009).
24. Lee, W. H. Binary computer-generated holograms. *Appl. Opt.* **18**, 3661–3669 (1979).
25. Lichte, H. & Lehmann, M. Electron holography - basics and applications. *Rep. Prog. Phys.* **71**, 016102 (2008).
26. Bliokh, K. Y., Dennis, M. R. & Nori, F. Relativistic electron vortex beams: angular momentum and spin-orbit interaction. *Phys. Rev. Lett.* **107**, 174802 (2011).
27. Bliokh, K. Y., Bliokh, Y. P., Savel'ev, S. & Nori, F. Semiclassical dynamics of electron wave packet states with phase vortices. *Phys. Rev. Lett.* **99**, 190404 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements N.V.-B. is an Eshkol scholar. This work was partly supported by the Israel Science Foundation and by the Israeli Ministry of Science. The author wishes to thank K. Shemer and D. Bloch for conversations.

Author Contributions N.V.-B. conceived the idea and designed the experiments. N.V.-B. and Y. Lereah carried out the experiment. Y. Lilach optimized the production process and fabricated the nanoscale holograms. A.G. and A.A. did the theoretical work and conceived ideas for applications. A.A. and N.V.-B. analysed the experimental results. A.G., A.A. and Y. Lereah provided experimental and theoretical guidance. A.A. and N.V.-B. developed and analysed the Airy lattice. All authors took part in writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.V.-B. (noavoloch@gmail.com).

Dicalcium nitride as a two-dimensional electride with an anionic electron layer

Kimoon Lee¹, Sung Wng Kim^{1,2}, Yoshitake Toda¹, Satoru Matsuishi³ & Hideo Hosono^{1,3}

Recent studies suggest that electrides—ionic crystals in which electrons serve as anions—are not exceptional materials but rather a generalized form, particularly under high pressure^{1–3}. The topology of the cavities confining anionic electrons determines their physical properties⁴. At present, reported confining sites consist only of zero-dimensional cavities or weakly linked channels⁴. Here we report a layered-structure electride of dicalcium nitride, Ca₂N, which possesses two-dimensionally confined anionic electrons whose concentration agrees well with that for the chemical formula of [Ca₂N]⁺·e[−]. Two-dimensional transport characteristics are demonstrated by a high electron mobility (520 cm² V^{−1} s^{−1}) and long mean scattering time (0.6 picoseconds) with a mean free path of 0.12 micrometres. The quadratic temperature dependence of the resistivity up to 120 Kelvin indicates the presence of an electron–electron interaction. A striking anisotropic magnetoresistance behaviour with respect to the direction of magnetic field (negative for the field perpendicular to the conducting plane and positive for the field parallel to it) is observed, confirming diffusive two-dimensional transport in dense electron layers. Additionally, band calculations support confinement of anionic electrons within the interlayer space, and photoemission measurements confirm anisotropic low work functions of 3.5 and 2.6 electronvolts, revealing the loosely bound nature of the anionic electrons. We conclude that Ca₂N is a two-dimensional electride in terms of [Ca₂N]⁺·e[−].

An electride, which possesses an electron as its anion, may be regarded as a crystalline form of solvated electrons^{4,5}. Because the physical properties derived from loosely bound anionic electrons should markedly differ from those of nearly free electrons in a metal or trapped electrons at specific sites such as F-centres, electrides could be useful electro-active materials^{4,6}. Until now, the topologies of cavities serving as sites to confine anionic electrons in electrides have been limited to zero-dimensional states or one-dimensionally linked channels; these restrictions inevitably mean that it is difficult for the anionic electrons in electrides to be delocalized^{4–6}.

We consider a two-dimensional (2D) quantum well state with which to confine anionic electrons in electrides. Two-dimensionally confined electrons have been realized by artificially fabricating hetero-interface structures such as GaAs/AlGaAs (ref. 7) or LaAlO₃/SrTiO₃ (ref. 8). Although it is rarely observed in bulk materials, to realize an electride with 2D anionic electrons a built-in 2D electron layer with a stoichiometric structure is required. To achieve built-in 2D anion electronic states, the candidate host material must meet three characteristics. First, the material must have a layered structure consisting of positively charged layer units, each of which act as a cationic slab. Second, the material must have an interlayer spacing that ensures the electrons are stably confined at room temperature (25 °C). Third, the material must have a simple component with a stable valence of metal cations to confine the loosely bound electron layers without trapping the electrons. Layered sub-nitrides meet these three requirements with respect to their crystal structure, interlayer spacing, and elements constituting the positively charged layer, [AE₂N]⁺ (where AE is an alkaline

earth metal)^{9,10}. To characterize AE₂N as an electride in terms of [AE₂N]⁺·e[−], it is critical that the number of anionic electron should agree with that expected from the chemical formula of [AE₂N]⁺·e[−], and that the anionic electrons should occur at interstitial positions (that is, the electrons must be primarily confined within the interlayer space, between the [AE₂N]⁺ layers).

Here we report that dicalcium nitride, Ca₂N, comprises a new class of electrides with 2D delocalized anionic electron layers. To investigate the 2D delocalized nature of the anionic electrons, we grew a single crystal of Ca₂N and measured its physical properties such as electron transport parameters and work function. Combined with the electronic structure calculation, all experimental results obtained indicate that Ca₂N is a 2D electride with a formula of [Ca₂N]⁺·e[−].

Figure 1a schematically depicts the crystalline structure of Ca₂N obtained by the Rietveld refinement of the powder X-ray diffraction pattern (Supplementary Fig. 1). As summarized in Supplementary Table 1, Ca₂N has a hexagonal layered structure in the *R*3*m* space group (anti-CdCl₂-type) with a high *c/a* ratio of 5.27, where *c* and *a* are unit cell dimensions. This *c/a* ratio is consistent with the values found in the literature^{9,10}. The layered unit is composed of edge-sharing NCa₆ octahedra in the [Ca₂N] layer, and each layer unit has a separation of 3.86 Å, which is about 50% larger than its thickness (2.51 Å). Because the distance between Ca and N (2.44 Å) is almost identical to the sum of the ionic radii of Ca²⁺ (1.0 Å) and N^{3−} (1.46 Å), the [Ca₂N] layer unit can be regarded as a positively charged ionic slab, [(Ca²⁺)₂N^{3−}]⁺. Thus, to compensate for the positive charge of the [Ca₂N]⁺ layers, the interlayer spacing should work as a confinement space for the anionic electrons, resulting in a [Ca₂N]⁺·e[−] configuration (Fig. 1a).

Figure 1b shows that Ca₂N forms a shiny green plate-like crystal, which exclusively exhibits reflection peaks from the (00*l*) plane, as is confirmed by the X-ray diffraction pattern. This result indicates that the shiny crystal surface is parallel to the (003) plane, which is where the 2D space is located. The crystal clearly grows faster along the plane perpendicular to the *c*-axis than along the plane parallel to the *c*-axis. This marked anisotropy is consistent with the results of mechanical exfoliation tests using 3M Scotch tape as reported in other layered-structure materials with highly anisotropic bonding natures¹¹. As shown in the inset of Fig. 1b, simple exfoliation produces a well cleaved surface of the (00*l*) plane.

Given its easily cleaved nature, we performed an optical reflectance measurement to evaluate the optical properties of Ca₂N. Figure 1c shows a sharp decrease in reflectance with photon energy and broad peaks occur at 1.5 eV and 2.4 eV. The green colour of the cleaved surface is correlated with the presence of the 2.4-eV peak. Hence, we assume that the plasma response due to the free carriers is responsible for the sharp decrease around 1.5 eV. We analysed this optical response quantitatively below using the information obtained from the Hall effect and electronic heat capacity measurements.

Figure 2a schematically depicts the stamping method of measuring the electrical properties of a Ca₂N single crystal, which allow reproducible

¹Frontier Research Center, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan. ²Department of Energy Science, Sungkyunkwan University, 300 Cheoncheon-dong, Jangan-gu, Suwon, Gyeonggi-do 440-746, South Korea. ³Materials and Structures Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan.

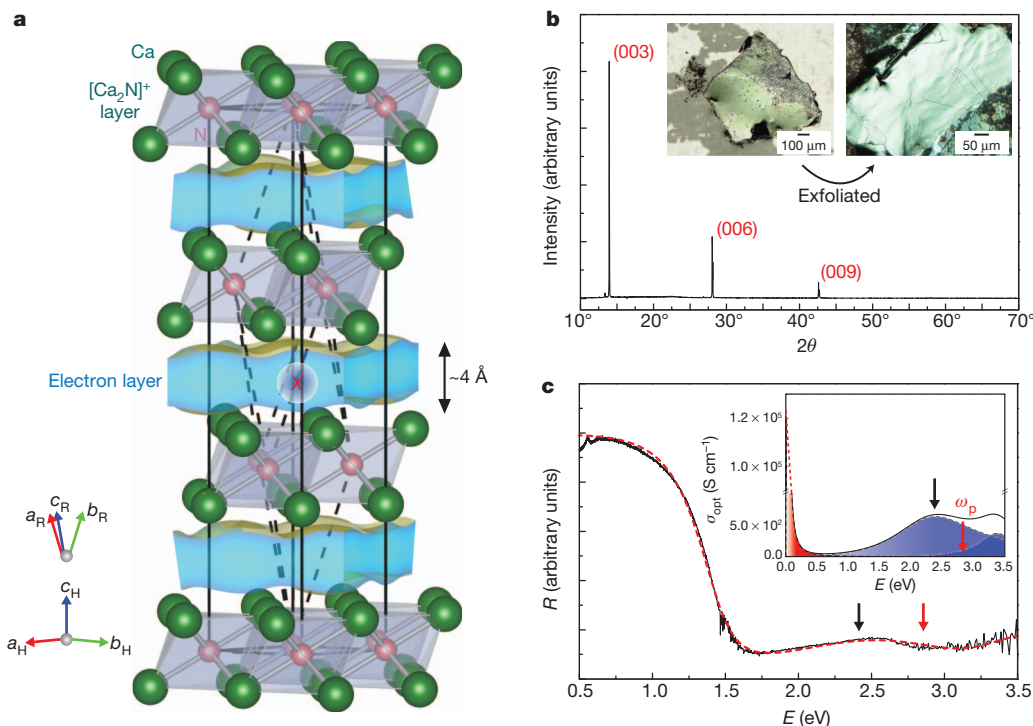


Figure 1 | Structural and optical characterizations of Ca_2N . **a**, Solid and dashed black lines represent the conventional hexagonal unit cell ($a_{\text{H}} = b_{\text{H}} \approx 3.6 \text{ \AA}$, $c_{\text{H}} \approx 19.1 \text{ \AA}$) and the primitive rhombohedral unit cell ($a_{\text{R}} = b_{\text{R}} = c_{\text{R}} \approx 6.7 \text{ \AA}$, $\alpha = \beta = \gamma \approx 31.0^\circ$), respectively. The interstitial site (red X) between the $[\text{Ca}_2\text{N}]^+$ layers is located at the body-centre position of the primitive cell. **b**, X-ray diffraction pattern measured by Cu $K\alpha$ radiation for a Ca_2N single crystal. Peaks from planes perpendicular to the hexagonal c -axis (c_{H}) are observed. Inset photos display the single-crystal surfaces before (left)

and after (right) exfoliation. **c**, Optical reflectance (R) spectrum of the cleaved single-crystal surface. Black and dashed red curves denote the observed and calculated traces based on the Drude–Lorentz model, respectively. The inset is the optical conductivity (σ_{opt}) spectrum calculated from the reflectance spectrum. The calculated σ_{opt} spectrum can be resolved into one Drude (red-shaded) and two Lorentz (blue-shaded) components. Positions corresponding to the plasma frequency ω_{p} and the point at which the sample turns green are indicated by red and black arrows, respectively.

data to be obtained¹². Figure 2b displays the temperature (T) dependence of resistivity (ρ) for both a polycrystalline pellet and a single crystal of Ca_2N . The single-crystal sample exhibits metallic transport with resistivity $\rho = 2.8 \mu\Omega \text{ cm}$ at 300 K, which is smaller than that of pure Ca metal ($3.6 \mu\Omega \text{ cm}$) (ref. 13). Furthermore, the quadratic T dependence is observed up to 120 K, indicating that the electron–electron interaction is much stronger than the electron–phonon interaction even in the high- T region¹³. Hall-effect measurements reveal that the primary charge carrier is an electron and the carrier concentration (n_{e}) at 300 K is $1.39 \times 10^{22} \text{ cm}^{-3}$, which agrees well with the calculated value ($1.37 \times 10^{22} \text{ cm}^{-3}$) based on three electrons per unit cell with a chemical formula of $[\text{Ca}_2\text{N}]^+ \cdot \text{e}^-$. Because n_{e} depends on T in the polycrystalline sample but not the single crystal (Fig. 2c), the observed T dependence is associated with the grain boundary nature. These results indicate that the large n_{e} value originates not from impurity doping, but from the intrinsic nature of the crystal^{13,14}.

As shown in Fig. 2d, μ_{H} increases from $160 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ to $520 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ as T decreases from 300 K to 2 K for the single-crystal sample. This value is superior to those of typical metals (several tens of $\text{cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) that can be described by a nearly free electron model¹⁴. Electronic heat capacity (C) measurements combined with Hall-effect measurements provide detailed information about the transport characteristics. As shown in the inset of Fig. 2c, a linear relationship exists between CT^{-1} and T^2 . The intercept and slope indicate that the Sommerfeld coefficient (γ) is $4.2 \text{ mJ mol}^{-1} \text{ K}^{-2}$ and the Debye temperature (Θ_{D}) is 228 K, respectively¹³. By comparing the experimental and theoretical values of γ , the effective mass (m^*) is calculated as $1.9m_{\text{e}}$ (ref. 13) (where m_{e} is the electron mass in free space). This m^* value is about twice that of the $[\text{Ca}_{24}\text{Al}_{28}\text{O}_{64}]^{4+} \cdot 4\text{e}^-$ (C12A7) electride¹⁵.

From the calculated optical conductivity (σ_{opt}) spectrum (see the inset of Fig. 1c) based on the Drude–Lorentz model^{15,16}, we can resolve

the Drude component with the plasma frequency (ω_{p} , denoted by a red arrow) at 2.8 eV. The obtained ω_{p} also enables us to calculate m^* as $2.5m_{\text{e}}$, which is similar to the value of $1.9m_{\text{e}}$ evaluated from the heat capacity measurements (see the Supplementary Information for a detailed analysis). The measured μ_{H} and m^* values allow mean scattering time (τ) and mean free path (l) to be calculated as 0.6 ps and 0.12 μm , respectively, based on the Drude model^{13,14,16}. These large τ and l values are comparable to those from typical 2D electronic systems rather than conventional metals (Supplementary Table 3)¹⁶. This comparison means that the electrons mainly conduct through a 2D space instead of the $[\text{Ca}_2\text{N}]^+$ layer, as depicted in Fig. 1a. The lack of a significant temperature dependence of μ_{H} in the 2–50 K range also suggests that the conduction electron is isolated from the source of scattering centres under minimal electron–phonon scattering.

To clarify the 2D transport behaviour of Ca_2N , we examined the anisotropy of resistivity under different magnetic field (H) directions. As shown in Fig. 2e–g, T -dependent negative magnetoresistance was observed under H perpendicular to the (00 l) plane of the crystal, whereas under in-plane H the magnetoresistance was conventionally parabolic and positive¹⁷. We note that negative magnetoresistance has been reported as a representative phenomenon arising mainly from the weak localization and/or electron–electron interaction in 2D weak disordered system ($l \gg a$, where a is the average atomic separation)¹⁸. When H is applied perpendicular to the conducting plane, the cyclotron orbit of the electron lies on the conducting plane [(00 l) for Ca_2N], and this movement suppresses the weak localization and/or electron–electron scattering, resulting in the decrease of resistivity. On the other hand, the cyclotron orbit for the in-plane H lies on the plane perpendicular to the (00 l) plane, which should sweep through the $[\text{Ca}_2\text{N}]^+$ slab with enhanced m^* (when $H \parallel a_{\text{H}}$, as in Supplementary Fig. 2). This means that it is hard to suppress the weak localization and/or electron

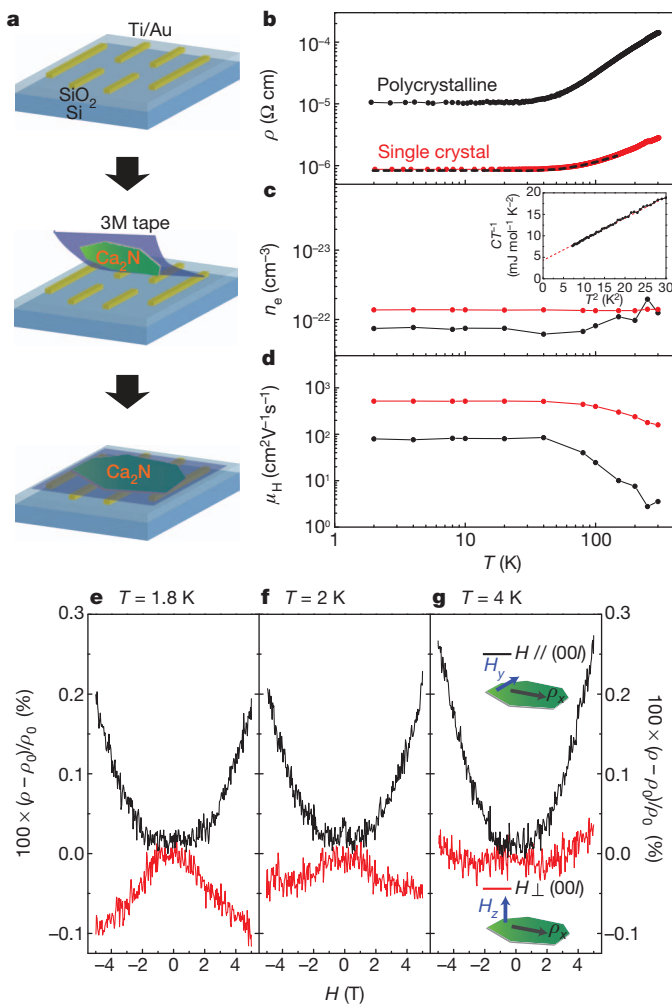


Figure 2 | Electron transport properties of Ca₂N single crystal. **a**, Schematic of sample preparation for a thin and freshly cleaved single crystal in order to attain a stable contact onto the electrodes. The Ti (5 nm)/Au (50 nm) electrodes were patterned by photolithography and deposited on a thermally grown SiO₂/Si substrate. After the single crystal was attached to a conventional 3M tape, the sample was aligned and pressed onto the patterned electrodes. **b–d**, Measured characteristics of resistivity (ρ), carrier concentration (n_e), and Hall mobility (μ_H), respectively, as functions of temperature (T) for a polycrystalline sample and single crystal. The dashed line indicates the quadratic T dependence of ρ . The inset to Fig. 2c shows the CT^{-1} versus T^2 plot from the heat capacity (C) measurement. The linear fit (red dashed line) shows that the free electrons contribute to the heat capacity. **e–g**, Magnetoresistance behaviour under out-of-plane [$H \perp (00l)$] and in-plane [$H \parallel (00l)$] magnetic field (H) for a single-crystal sample. The insets give the schematic illustrations of the geometry for magnetoresistance measurements. As T decreases, magnetoresistance for out-of-plane H becomes negative relative to ρ_0 (ρ at $H = 0$), while magnetoresistance for in-plane H remains positive.

interactions using the in-plane H (ref. 18). These marked anisotropic magnetoresistance results confirm that the diffusive 2D transport in the dense electron layer along the (00 l) plane occurs as we expected. The measured γ value of $4.2 \text{ mJ mol}^{-1} \text{ K}^{-2}$ is much larger than that of pure Ca ($2.9 \text{ mJ mol}^{-1} \text{ K}^{-2}$) notwithstanding that its n_e ($1.39 \times 10^{22} \text{ cm}^{-3}$) is lower than that of Ca ($4.6 \times 10^{22} \text{ cm}^{-3}$; ref. 13). This means that electron–electron interaction in Ca₂N is much stronger than in Ca metal, suggesting that the electrons are delocalized not throughout the entire crystal, but in a lower-dimensional space. Each of the results obtained supports the hypothesis that the transport characteristics of Ca₂N result from a well defined 2D electronic state.

Figure 3a and b display the calculated band structure and density of states profiles based on the density functional theory (DFT), respectively.

The single band crossing the Fermi level (E_F) and forming a broad density-of-states peak around E_F is mainly composed of a non-atom-centred orbital located around X (see Fig. 1a) with small contributions from the atomic orbitals of Ca and N, as shown in projected density-of-states results. Figure 3c shows the band-decomposed electron density maps for the ‘interstitial band’ on the (1 $\bar{1}$ 0)_R plane of rhombohedral representation. In particular, the conduction electrons occupying electronic states within an energy ($kT = 25 \text{ meV}$) corresponding to near E_F at room temperature possess a spherical distribution around X. This result indicates that the electrical conduction of Ca₂N is due to electrons confined within the interstitial space between [Ca₂N]⁺ layers as illustrated in Fig. 1a. The calculated electron-localization function shows the unshared-electron interactions between interstitial functions (at X) and the [Ca₂N]⁺ layer, reflecting ionic bonding (dashed white circles in Fig. 3c)¹⁹. The integrated electron density of the conduction band from -1.48 eV to 0.00 eV is equal to 1. This result is consistent with the chemical view of the stoichiometric electride (Ca²⁺)₂(N³⁻)⁺·e⁻ with an excess electron density of $1.37 \times 10^{22} \text{ cm}^{-3}$. We note that n_e ($1.39 \times 10^{22} \text{ cm}^{-3}$) obtained by Hall measurement also agrees with this theoretical value, supporting the same result. Given that these interstitial electrons constitute the cylindrical Fermi surface (Fig. 3d), like a typical 2D electronic system, we conclude that the conduction electrons at the Fermi surface primarily result from the 2D confined anionic electrons. The Fermi surface information obtained also provides m^* of the conduction electrons ($1.8\text{--}2.2m_e$) for the ab -plane (see Supplementary Fig. 2; ref. 20), which is consistent with experimental values ($1.9\text{--}2.5m_e$). All results obtained by DFT calculations thus strongly support the experimental result that Ca₂N is a 2D electride with the formula [Ca₂N]⁺·e⁻ in which electrons function as counter anions and exhibit 2D transport.

Finally, to reveal the loosely bound character of the conduction electrons, we measured the work function by ultraviolet photoelectron spectroscopy (UPS). Although UPS measurements have been reported for various alkaline earth metal sub-nitrides, the work function has not been successfully measured, probably owing to difficulties preparing a clean surface²¹. To overcome these difficulties, we tried to cut the polycrystalline sample or exfoliate a single crystal under ultrahigh-vacuum conditions using a homemade device (Fig. 4a). As shown in Fig. 4b and c, the cut-off energy value linearly shifts to a lower kinetic energy as the applied bias (V_{bias}) increases for the single crystal sample, but the cut-off energy is nearly constant for the polycrystalline sample. When we checked the valence band spectra of the single crystal sample, V_{bias} gives rise to the broadening of valence band (inset of Fig. 4b). Additionally, the inset of Fig. 4c demonstrates that the magnitude of the cut-off energy shift depends on the photon yield. These results indicate that the shift in the cut-off energy in the single crystal originates mainly from the band bending, which is induced by the surface photo-voltage effect in the absence of free carriers at the surface even though the single crystal clearly shows a metallic nature with a Fermi edge, as denoted by E_F (ref. 22). The cut-off energy of the secondary electron emission is mainly determined by the surface consisting of the [Ca₂N]⁺ slab at (00 l) plane, so we infer that the [Ca₂N]⁺ single slab is electrically resistive to induce such a photo-voltage effect even though conducting electrons are beneath the slab. According to the DFT calculations, the anionic electron layer should occupy the interstitial space beneath the slab, and not belong to the atomic orbitals in the [Ca₂N]⁺ slab. Thus, we conclude that the bias voltage (V_{bias})-independent cut-off shown in Fig. 4c is due to other conductive planes, which can directly expose anionic electrons, such as the (100) one.

The inset in Fig. 4c shows the values of the work functions, which were determined by extrapolating the cut-off energy values to $V_{\text{bias}} = 0$ (ref. 23). The work function for a single crystal with a (003)-related plane (‘related’ here means any other planes parallel to (003), such as (001) or (002)) occurs at 3.5 eV , which is about 1 eV higher than that for the polycrystalline sample (2.6 eV). It is noteworthy that a first-principles calculation also predicts an anisotropic work function for

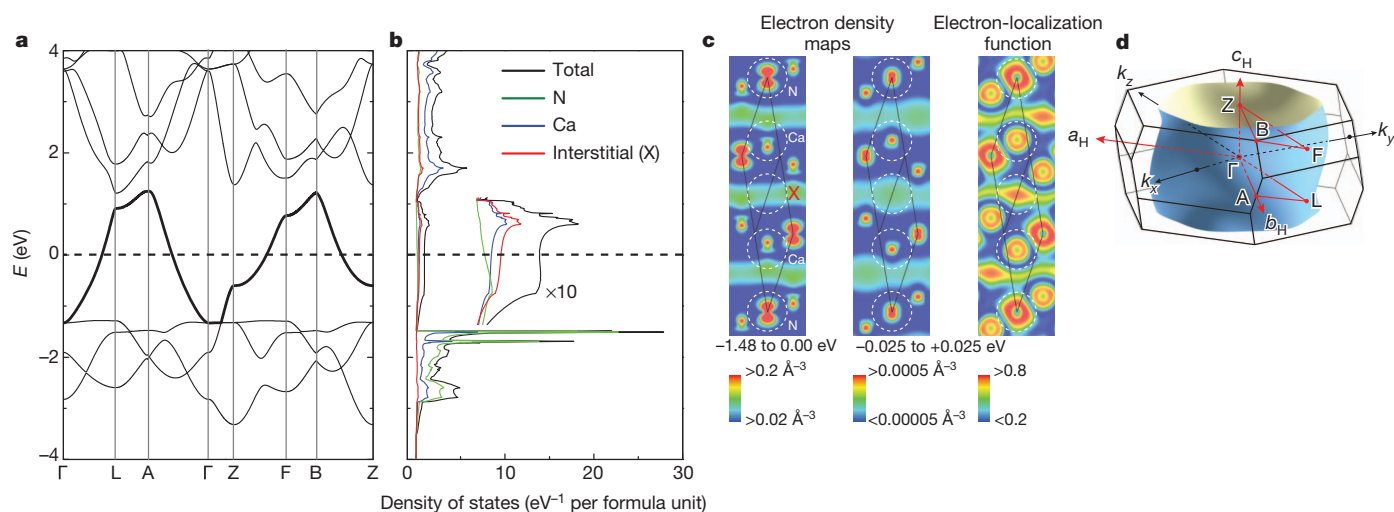


Figure 3 | Calculated electronic structure of Ca_2N . **a**, Band structure. E_F is taken as the origin (dashed line), and the bold dispersion curve crossing E_F represents the 'interstitial band', which is mainly composed of electronic states located around X, as shown in Fig. 1a. **b**, Total and projected density of states on the spheres located at Ca, N and X positions. **c**, Electron density maps of interstitial band (left and centre panels) and the electron-localization function

of the total electron density (right panel) on the $(1\bar{1}0)_R$ plane parallel to the hexagonal c -axis (c_H), bisecting a rhombohedral unit cell. The left panel shows the electron density for the total occupied states of the interstitial band ($E = -1.48$ eV to 0.00 eV) and the centre panel shows the electron density for the room-temperature (kT) region near E_F ($|E| < 0.025$ eV). **d**, The Fermi surface in the first Brillouin zone.

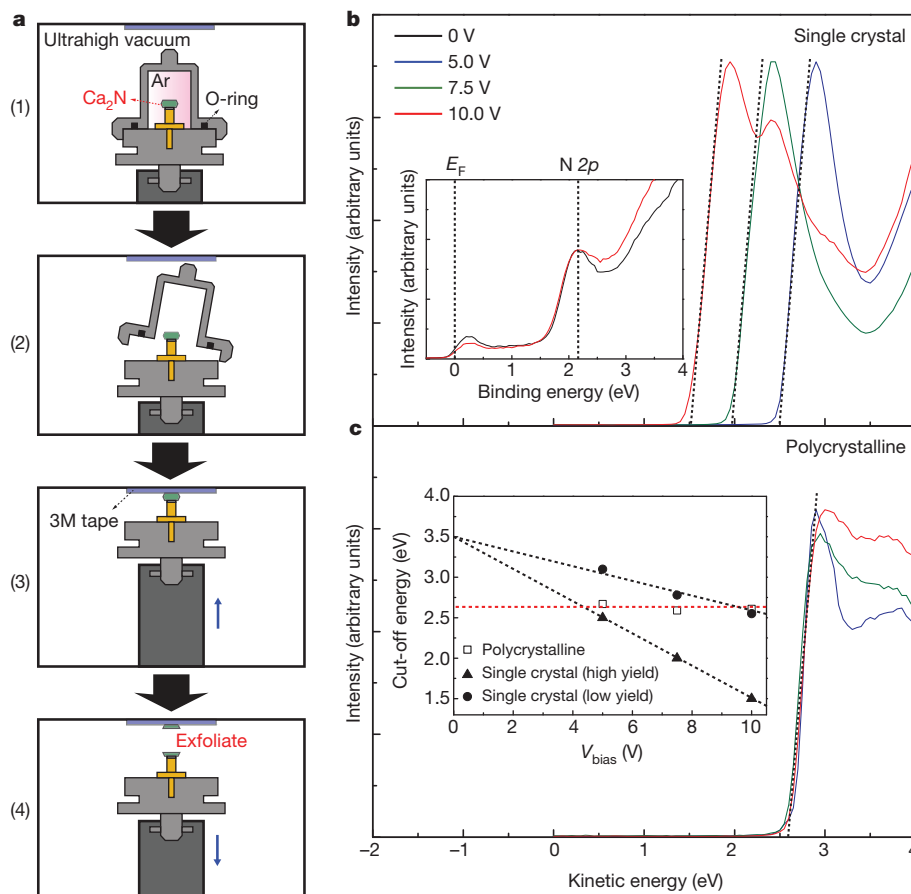


Figure 4 | Determination of the work function by UPS. **a**, Procedure to obtain a clean surface of a single crystal under *in situ* conditions in an ultrahigh-vacuum chamber. (1) The sample, which was capped in the O-ring sealed stage under Ar, was transferred into the ultrahigh-vacuum chamber. (2) After evacuating the chamber, the cap was opened. (3) The stage was moved so it could be attached to 3M tape. (4) Lowering the stage exfoliated the sample under ultrahigh vacuum. **b**, **c**, Secondary electron emission from the single

crystal and polycrystalline pellet surfaces, respectively. The inset of **b** shows the valence band spectra of the (001) surfaces under V_{bias} of 0 V (black line) and 10 V (red line). Each sample has a distinct Fermi edge for various V_{bias} values, and by extrapolating the cut-off energies to $V_{\text{bias}} = 0$ we determined the work function values (inset of **c**). Fermi edge (E_F) location demonstrates the metallic nature of sample.

Ca_2N where the (100) plane is less than that for the (001) plane²⁴. Furthermore, the values obtained here are close to the previous calculated ones—3.43 eV for (001) and 2.56 eV for the un-relaxed (100) plane²⁴—although the measured values could be modified by the surface reconstruction, especially for the (001) plane²⁵ (refer to the Supplementary Information). UPS on the polycrystalline sample gives the lowest work function because the secondary electron cut-off due to the higher work function is effectively screened by the cut-off from the lowest one²⁶. Therefore, we conclude that the lowest work function at 2.6 eV for the polycrystalline sample results from the (100) related plane. It should be noted that the work function of Ca_2N (100) is smaller by about 0.3 eV than the value of metal Ca (2.9 eV); such a low work function of a cathode of an organic light-emitting diode would help to reduce the electron injection barrier.

We examined the physical properties of single crystalline Ca_2N , and discovered that it can be expressed as $[\text{Ca}_2\text{N}]^+ \cdot \text{e}^-$. Distinct 2D electronic properties can be explained as a result of the built-in electron layers in the 2D defined space, arousing the itinerant anionic state (that is, the anionic electrons are well delocalized). The present $[\text{Ca}_2\text{N}]^+ \cdot \text{e}^-$ is the most delocalized anionic electronic system among electride family reported to date. We anticipate that this 2D electride will stimulate the development of a series of 2D electrides, leading to practical applications, such as recently reported for the C12A7 electride²⁷.

METHODS SUMMARY

A stoichiometric polycrystalline Ca_2N sample was synthesized by the solid-state reaction method with Ca_3N_2 powder and Ca shots in a 1:1 ratio. The single crystals of Ca_2N were grown by the conventional self-flux method using Ca as a reactive flux metal. To measure the electrical properties of a polycrystalline pellet ($2 \times 1 \times 0.1 \text{ mm}^3$), the electrical contacts in the four-probe and Hall bar configurations were made by applying Ag pastes onto a sample. For the smaller single-crystal sample ($0.5 \times 0.5 \times 0.02 \text{ mm}^3$), we adopted the stamp method, which has been used to measure the electrical properties of organic single crystals¹². As shown in Fig. 2a, a plate-like single crystal of Ca_2N was attached to 3M Scotch tape and then the sample was pressed onto the lithographically patterned electrodes on a Si/SiO₂ wafer. DFT calculations were performed using the generalized gradient approximation Perdew–Burke–Ernzerhof functional, and the projected augmented plane-wave method^{28,29} implemented in the Vienna *ab initio* simulation program code³⁰. UPS measurements were carried out by using a hemispherical analyser with a He I light source. To prepare a clean Ca_2N surface free from surface contamination, we broke the polycrystalline pellet under ultrahigh vacuum. For the single crystal, the sample was also attached by conductive epoxy on the specially designed holder to seal the sample during transfer (Fig. 4a). After the vacuum level reached about 10^{-4} Pa in the chamber, which included the holder, the cap was opened and the single-crystal sample was exfoliated by 3M tape mounted beforehand. This method resulted in a clear cleaved surface. All manipulations were performed in a purified Ar-filled glove box because Ca_2N is highly reactive with ambient oxygen and water vapour.

Full Methods and any associated references are available in the online version of the paper.

Received 17 June; accepted 20 November 2012.

Published online 30 January 2013.

- Pickard, C. J. & Needs, R. J. Dense low-coordination phase of lithium. *Phys. Rev. Lett.* **102**, 146401 (2009).
- Ma, Y. *et al.* Transparent dense sodium. *Nature* **458**, 182–185 (2009).
- Pickard, C. J. & Needs, R. J. Aluminium at terapascal pressures. *Nature Mater.* **9**, 624–627 (2010).

- Dye, J. L. Electrides: early examples of quantum confinement. *Acc. Chem. Res.* **42**, 1564–1572 (2009).
- Kim, S. W., Shimoyama, T. & Hosono, H. Solvated electrons in high-temperature melts and glasses of the room-temperature stable electride $[\text{Ca}_{24}\text{Al}_{28}\text{O}_{64}]^{4+} \cdot 4\text{e}^-$. *Science* **333**, 71–74 (2011).
- Toda, Y. *et al.* Field emission of electron anions clathrated in subnanometer-sized cages of $[\text{Ca}_{24}\text{Al}_{28}\text{O}_{64}]^{4+} (4\text{e}^-)$. *Adv. Mater.* **16**, 685–689 (2004).
- Ando, T., Fowler, A. B. & Stern, F. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.* **54**, 437–672 (1982).
- Ohtomo, A. & Hwang, H. Y. A high mobility electron gas at the $\text{LaAlO}_3/\text{SrTiO}_3$ heterointerface. *Nature* **427**, 423–426 (2004).
- Gregory, D. H., Bowman, A., Baker, C. F. & Weston, D. P. Dicalcium nitride, Ca_2N -a 2D “excess electron” compound; synthetic routes and crystal chemistry. *J. Mater. Chem.* **10**, 1635–1641 (2000).
- Reckeweg, O. & DiSalvo, F. J. Alkaline earth metal nitride compounds with the composition M_2NX ($\text{M} = \text{Ca, Sr, Ba}$; $\text{X} = \square, \text{H, Cl or Br}$). *Solid State Sci.* **4**, 575–584 (2002).
- Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
- Sundar, V. C. *et al.* Elastomeric transistor stamps: reversible probing of charge transport in organic crystals. *Science* **303**, 1644–1646 (2004).
- Kittel, C. *Introduction to Solid State Physics* 8th edn (John Wiley & Sons, 2005).
- Kasap, S. O. *Principles of Electronic Materials and Devices* 2nd edn (McGraw-Hill, 2002).
- Matsuishi, S. *et al.* Localized and delocalized electrons in room-temperature stable electride $[\text{Ca}_{24}\text{Al}_{28}\text{O}_{64}]^{4+} (\text{O}^{2-})_{2 \times 2} (\text{e}^-)_{2 \times 2}$: analysis of optical reflectance spectra. *J. Phys. Chem. C* **112**, 4753–4760 (2008).
- Fox, M. *Optical Properties of Solids* (Oxford Univ. Press, 2001).
- Oslen, J. L. *Electron Transport in Metals* (Interscience, 1962).
- Pirau, L. Weak localization and coulomb interaction in graphite intercalation compounds and related materials. *J. Mater. Res.* **5**, 1285–1298 (1990).
- Li, Z., Yang, J., Hou, J. G. & Zhu, Q. Is mayenite without clathrated oxygen an inorganic electride? *Angew. Chem. Int. Edn* **43**, 6479–6482 (2004).
- Rourke, P. M. C. & Julian, S. R. Numerical extraction of de Haas–van Alphen frequencies from calculated band energies. *Comput. Phys. Commun.* **183**, 324–332 (2012).
- Steinbrenner, U., Adler, P., Hölle, W. & Simon, A. Electronic structure and chemical bonding in alkaline earth metal subnitrides: photoemission studies and band structure calculations. *J. Phys. Chem. Solids* **59**, 1527–1536 (1998).
- Demuth, J. E., Thompson, W. J., DiNardo, N. J. & Imbühl, R. Photoemission-based photovoltage probe of semiconductor surface and interface electronic structure. *Phys. Rev. Lett.* **56**, 1408–1411 (1986).
- Toda, Y. *et al.* Work function of a room-temperature, stable electride $[\text{Ca}_{24}\text{Al}_{28}\text{O}_{64}]^{4+} (\text{e}^-)_4$. *Adv. Mater.* **19**, 3564–3569 (2007).
- Uijtewaald, M. A. de Wijs, G. A. & de Groot, R. A. Low work function of the (1000) Ca_2N surface. *J. Appl. Phys.* **96**, 1751–1753 (2004).
- Tasker, P. W. The stability of ionic crystal surfaces. *J. Phys. C* **12**, 4977–4984 (1979).
- Grepstad, J. K., Gartland, P. O. & Slagsvold, B. J. Anisotropic work function of clean and smooth low-index faces of aluminium. *Surf. Sci.* **57**, 348–362 (1976).
- Kitano, M. *et al.* Ammonia synthesis using a stable electride as an electron donor and reversible hydrogen store. *Nature Chem.* **4**, 934–940 (2012).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
- Kresse, G. & Furthmüller, J. Efficient iterative scheme for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST), JSPS, and the Element Strategy Initiative Project, MEXT, Japan.

Author Contributions H.H. conceived and S.W.K. initiated the study. K.L. and S.W.K. synthesized the samples and measured electron transport properties. Y.T. carried out the UPS measurement. S.M. performed DFT calculations. K.L., S.W.K. and H.H. co-wrote the manuscript. All the authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.H. (hosono@msl.titech.ac.jp).

METHODS

All manipulations were performed in a purified Ar-filled glove box because Ca_2N is highly reactive with ambient oxygen and water vapour.

Sample synthesis. To synthesize a stoichiometric polycrystalline Ca_2N sample, we mixed Ca_3N_2 powder and Ca shots in a 1:1 ratio, and then pressed them into a pellet form. After the sample was wrapped with molybdenum foil, it was annealed at 800 °C for 48 h under vacuum ($\sim 10^{-3}$ Pa) by pumping with a vacuum pump connected to a silica tube in a box furnace. Then the sample was quenched into water. By employing this quenching process, we successfully eliminated the formation of secondary phases, except for the inevitable trace amount of CaO, by suppressing the potential intermediate reactions in the moderate temperature range. The synthesized sample was ground into a powder in an agate mortar and re-annealed under the same conditions to improve homogeneity.

A single crystal of Ca_2N was synthesized by the conventional self-flux method using Ca as a reactive flux metal. We pressed a mixture of Ca_3N_2 powder and Ca metal in a 1:10 ratio into a pellet, which was subsequently wrapped with molybdenum foil and sealed in a silica glass tube while degassing. After the sealed sample was annealed at 800 °C for 24 h, we slowly cooled the furnace to 600 °C at a rate of 0.67 °C per hour. Then the sample was cooled to room temperature by turning off the heat.

Electrical measurements. To measure the electrical properties of a polycrystalline pellet ($2 \times 1 \times 0.1 \text{ mm}^3$), the electrical contacts in the four-probe and Hall bar configurations were made by applying Ag pastes onto a sample. Apiezon N grease was lightly coated onto the sample surfaces after contact fabrication to prevent the sample from reacting with ambient air during the measurement. For the smaller single-crystal sample ($0.5 \times 0.5 \times 0.02 \text{ mm}^3$), we adopted the stamp method, which has been applied to measure the electrical properties of organic single crystals¹². As shown in Fig. 2a, a plate-like single crystal of Ca_2N was sustained

by 3M scotch tape and then the sample was pressed with 3M tape onto the lithographically patterned electrodes on a Si/SiO_2 wafer.

Electronic structure calculations. DFT calculations were performed using the generalized gradient approximation (GGA) Perdew–Burke–Ernzerhof functional, and the projected augmented plane-wave method^{28,29} implemented in the Vienna *ab initio* simulation program code³⁰. A primitive rhombohedral unit cell containing one chemical formula was used, and the plane-wave basis-set cut-off was set to 600 eV. To obtain the projected density of states, the charge density was decomposed over the atom-centred spherical harmonics with a Wigner–Seitz radius of $\sim 1.7 \text{ \AA}$. We set the ‘interstitial region’ as an empty sphere placed at the body-centre position of the unit cell as denoted by X in Fig. 3c. Fermi surfaces were calculated using WIEN2k code³¹ and m^* was calculated using SKEAF code²⁰. The value of m^* calculated corresponds to the cyclotron mass upon applying H in the c_H direction. More details are described in the Supplementary Information.

UPS measurement. Measurements were performed by using a hemispherical analyser with a He I light source. To prepare a clean Ca_2N surface free from surface contamination, we broke the polycrystalline pellet under ultrahigh vacuum. For the single crystal, the sample was also attached by conductive epoxy on the specially designed holder to seal the sample during transfer (Fig. 4a). After the vacuum level reached $\sim 10^{-4}$ Pa in the chamber, which included the holder, the cap was opened and the single-crystal sample was exfoliated by 3M tape mounted beforehand. This method resulted in a clear cleaved surface. During the UPS measurement, a He I beam illuminated the sample surface at a 45° angle to the surface normal, which resulted in a $\sim 2\text{-mm}$ -diameter beam spot. Each measurement was performed at room temperature under the base pressure of 5×10^{-8} Pa.

31. Blaha, P., Schwarz, K., Madsen, G. K. H., Kvasnicka, D. & Luitz, J. *An Augmented Plane and Local Orbitals Program for Calculating Crystal Properties* (ed. Schwarz, K) (Technical University of Wien, 2001).

Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability

Peter M. Cox¹, David Pearson², Ben B. Booth², Pierre Friedlingstein¹, Chris Huntingford³, Chris D. Jones² & Catherine M. Luke¹

The release of carbon from tropical forests may exacerbate future climate change¹, but the magnitude of the effect in climate models remains uncertain². Coupled climate–carbon-cycle models generally agree that carbon storage on land will increase as a result of the simultaneous enhancement of plant photosynthesis and water use efficiency under higher atmospheric CO₂ concentrations, but will decrease owing to higher soil and plant respiration rates associated with warming temperatures³. At present, the balance between these effects varies markedly among coupled climate–carbon-cycle models, leading to a range of 330 gigatonnes in the projected change in the amount of carbon stored on tropical land by 2100. Explanations for this large uncertainty include differences in the predicted change in rainfall in Amazonia^{4,5} and variations in the responses of alternative vegetation models to warming⁶. Here we identify an emergent linear relationship, across an ensemble of models⁷, between the sensitivity of tropical land carbon storage to warming and the sensitivity of the annual growth rate of atmospheric CO₂ to tropical temperature anomalies⁸. Combined with contemporary observations of atmospheric CO₂ concentration and tropical temperature, this relationship provides a tight constraint on the sensitivity of tropical land carbon to climate change. We estimate that over tropical land from latitude 30° north to 30° south, warming alone will release 53 ± 17 gigatonnes of carbon per kelvin. Compared with the unconstrained ensemble of climate–carbon-cycle projections, this indicates a much lower risk of Amazon forest dieback under CO₂-induced climate change if CO₂ fertilization effects are as large as suggested by current models⁹. Our study, however, also implies greater certainty that carbon will be lost from tropical land if warming arises from reductions in aerosols¹⁰ or increases in other greenhouse gases¹¹.

We use results from the Coupled Climate Carbon Cycle Model Intercomparison Project³ (C⁴MIP) focusing on changes in tropical land carbon storage in the latitudinal band from 30° N to 30° S. Although C⁴MIP included general circulation models (GCMs) and Earth-system models of intermediate complexity, we limit our analysis to the GCMs because our emergent constraint requires models that generate interannual variability. The C⁴MIP experimental design³ forced models using the SRES A2 scenario¹² of anthropogenic CO₂ emissions (including those due to land-use change). For each model, an ‘uncoupled’ simulation was carried out in which the land and ocean carbon cycles were made insensitive to the climate change caused by the increase in atmospheric CO₂. Comparison between the coupled and uncoupled simulations allows the direct effects of CO₂ on land and ocean carbon sinks to be separated from the effects of climate change^{3,13}. We test the emergent constraint derived from the C⁴MIP GCMs against results from the recent HadCM3 land carbon-cycle ensemble¹⁴.

Our emergent constraint could also be tested against the recent CMIP5 climate–carbon-cycle models, which will appear in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. However, models in that report typically use prescribed

concentrations of atmospheric CO₂ (ref. 15). This makes direct comparison with the observed interannual variability in the atmospheric CO₂ concentration difficult. Therefore, the emergent constraint we present here is conditional on the relatively simplistic representations of the carbon cycle in the C⁴MIP models.

Table 1 summarizes results from six C⁴MIP GCMs (A to F) for 1960 to 2099. For all models, the impact of climate change on the carbon cycle results in a larger increase in atmospheric CO₂ in the coupled simulation relative to the uncoupled simulation. This amplification varies by an order of magnitude across the model ensemble (from an extra 18 parts per million by volume (p.p.m.v.) in model D to an extra 212 p.p.m.v. in model A). A large part of this uncertainty arises from differing responses of tropical land carbon to projected climate changes in each model. All models produce a significant increase in tropical land carbon storage in the uncoupled simulations as a result of the direct effects of CO₂ on photosynthesis and water-use efficiency (from +263 gigatonnes of carbon (GtC) in model F to +413 GtC in model C). The neglect of carbon–nitrogen interactions in this first generation of climate–carbon models is arguably a major limitation in the mid and high latitudes¹⁶, but is much less problematic in tropical forests, which are not typically nitrogen-limited¹⁷. Forest inventories are also consistent with a significant CO₂ fertilization in the tropics^{18,19}. Despite the reasonable agreement among models on the effect of CO₂ fertilization, the fully coupled simulations produce very different changes in tropical land carbon storage from 1960 to 2099 (from –11 GtC for model A to +319 GtC for model D).

Figure 1a represents the evolution of tropical land carbon storage in the C⁴MIP models, with the upper and lower estimates shown for both the coupled and uncoupled simulations. The lower estimate in the coupled simulation comes from the HadCM3LC model, which projects Amazon forest dieback under CO₂-induced climate change^{19,10}. In this model, tropical land carbon storage increases owing to direct CO₂ effects until around 2050, but then declines abruptly owing to warming and drying in Amazonia⁹. This projection, along with recent extreme droughts in Amazonia^{20–22}, suggests that tropical forest dieback is a potential high-impact tipping element that would constitute an abrupt change in Earth’s climate system²³.

To separate direct effects of CO₂ from those of climate change, we follow previous analyses^{3,13} in writing the change in tropical land carbon storage, ΔC_{LT} , in terms of the change in atmospheric CO₂, ΔC_a , and the change in tropical mean temperature, ΔT_T :

$$\Delta C_{LT} = \beta_{LT} \Delta C_a + \gamma_{LT} \Delta T_T$$

Here β_{LT} (GtC per p.p.m.v.) and γ_{LT} (GtC K^{–1}) are the sensitivities of tropical land carbon storage to direct CO₂ effects and to climate change, respectively. The uncoupled simulations are used to estimate β_{LT} for each model, and then these values are used to isolate γ_{LT} from the coupled simulations^{3,13} by subtracting the direct CO₂ effect. Figure 1b is a scatter plot of β_{LT} and γ_{LT} for each C⁴MIP model and the three HadCM3 ensemble members. Whereas the β_{LT} values span a factor of two, from about 0.5 to 1 GtC per p.p.m.v., the γ_{LT} values range

¹College of Engineering, Mathematics and Physical Science, University of Exeter, Exeter EX4 4QF, UK. ²Hadley Centre, Met Office, Exeter EX1 3PB, UK. ³Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK.

Table 1 | Summary data for climate-carbon cycle projections

| Model | | Change in global atmospheric CO ₂ (p.p.m.v.) | | Change in tropical land carbon (GtC) | | Change in tropical temperature (K) |
|-------|------------|---|-----------|--------------------------------------|-----------|------------------------------------|
| | | Coupled | Uncoupled | Coupled | Uncoupled | |
| A | HadCM3LC | 689 | 477 | −11 | 354 | 3.93 |
| B | IPSL | 453 | 381 | 177 | 365 | 2.70 |
| C | MPI | 524 | 443 | 242 | 413 | 4.36 |
| D | CCSM1 | 483 | 465 | 319 | 364 | 1.53 |
| E | FRCGC | 589 | 465 | 118 | 271 | 3.61 |
| F | LOOP | 489 | 460 | 185 | 263 | 3.30 |
| G | HadCM3C-st | 599 | 331 | −148 | 317 | 4.41 |
| H | HadCM3C-a | 445 | 333 | −6 | 168 | 3.76 |
| I | HadCM3C-h | 589 | 246 | −165 | 251 | 4.08 |

Changes in atmospheric CO₂, tropical land carbon and tropical near-surface air temperature (30° N–30° S), as simulated by the nine climate–carbon GCMs analysed in this study. Models A to F are from the C⁴MIP study³, which prescribed the SRES A2 CO₂ emissions scenario. For these models, the changes are calculated over the period 1960–2099. Models G to I are from a land carbon-cycle parameter ensemble carried out with the HadCM3 model under the SRES A1B scenario¹⁴, and were run only to 2080, so differences here are for 1960 to 2080. In all cases, model runs were carried out both including and excluding climate effects on the carbon cycle ('coupled' and 'uncoupled', respectively), so that the impacts of climate–carbon-cycle feedbacks could be diagnosed.

over a factor of more than four, from -29 GtC K^{-1} (model F) to -133 GtC K^{-1} (model A), with a C⁴MIP mean of -69 GtC K^{-1} and standard deviation of 39 GtC K^{-1} . This range is even larger if the HadCM3 ensemble members are included. We therefore focus on reducing the larger uncertainty, namely that in γ_{LT} .

Our inspiration for deriving a multi-model emergent constraint comes from a recent study that showed a strong relationship between the contemporary temperature sensitivity of seasonal snow cover and the magnitude of the snow–albedo feedback, across more than 20 GCMs⁷. Because the seasonal cycle of snow cover can be estimated from observations, this model-derived relationship converts the contemporary observations to a constraint on the size of the snow–albedo feedback in the real climate system, for which there is no direct reliable measurement. Emergent constraints of this type make use of the often bewildering spread among Earth-system model projections to reduce uncertainties in the sensitivities of the real Earth system to anthropogenic forcing. They are distinct and complementary to bottom-up constraints arising from process-based studies.

It made sense a priori to look for an emergent constraint linking the sensitivity of tropical land carbon to interannual variability (IAV) in the growth rate of atmospheric CO₂. Tropical land carbon changes in response to climate through changes in the net land–atmosphere CO₂ flux into and out of this carbon store. Critically, the sensitivity of this net tropical CO₂ flux is revealed by the IAV in the CO₂ growth rate, because this is known to be dominated by the response of the tropical land carbon cycle to climatic anomalies (Supplementary Fig. 1a) such as the El Niño/Southern Oscillation^{8,24,25}. Hence, some relationship between the IAV in CO₂ and the longer-term sensitivity of tropical land carbon storage to climate change (γ_{LT}) is to be expected, as long as processes that are not evident in the short-term variation of the CO₂ fluxes (for example forest dynamics or changes in long-lived soil carbon pools) do not dominate the long-term response. This is our

working hypothesis to be tested against the C⁴MIP models, which include a range of representations of slow vegetation and soil processes³.

Figure 2a compares the observed IAV in the growth rate of global atmospheric CO₂ (refs 26, 27) with the IAV in the annual mean tropical temperature²⁸. In both cases, we have chosen observational variables (global mean atmospheric CO₂ and mean land-plus-ocean temperature between 30° N and 30° S) for consistency with the variables available from the C⁴MIP models. Aside from the years immediately after the volcanic eruptions²⁴ of Mount Agung, El Chichon and Mount Pinatubo, the IAV in the growth rate of atmospheric CO₂ is linearly correlated with the IAV in the tropical temperature ($r = 0.65$ (correlation coefficient), $P < 0.0001$; Fig. 2b), with a best-fit 'IAV sensitivity' of $5.1 \pm 0.9 \text{ GtC yr}^{-1} \text{ K}^{-1}$. Excluding these volcano-affected years has an impact on the best-fit sensitivity of less than 5%, but avoids the complication of diffuse-light fertilization of plant growth²⁹, which is not included in any of the C⁴MIP models. We also find a similar sensitivity regardless of which tropical temperature reconstruction we use. There is a greater sensitivity to the choice of the global atmospheric CO₂ data set, but this does not affect our overall conclusions (Supplementary Table 1).

A similar calculation is made for each of the coupled climate–carbon-cycle models, to derive the sensitivity of the CO₂ growth rate to tropical temperature for the period 1960–2010. Compared with the observational data, models tend to overestimate the IAV in the tropical temperature by a factor of up to two, and to overestimate the IAV in the CO₂ growth rate by a factor of up to three. The correlation between these variables is underestimated in some models (F, B and D) and overestimated in others (A, E and C). Hence, IAV sensitivity varies across the C⁴MIP model ensemble, from $2.9 \pm 1.4 \text{ GtC yr}^{-1} \text{ K}^{-1}$ (model F) to $9.7 \pm 0.7 \text{ GtC yr}^{-1} \text{ K}^{-1}$ (model A), with most of this range resulting from differences in the sensitivity of heterotrophic respiration to climate (Supplementary Fig. 1b). The three HadCM3

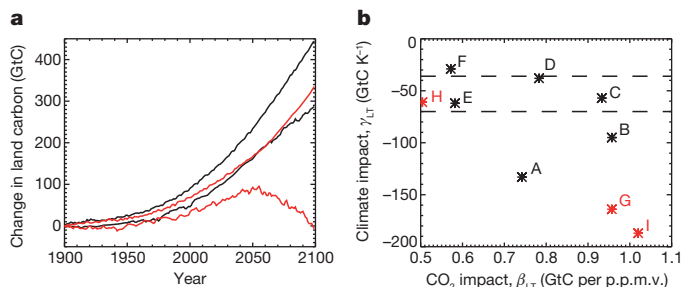


Figure 1 | Projected changes in land carbon storage in the tropics from coupled climate-carbon-cycle models. **a**, Upper and lower estimates from the C⁴MIP models³ (A–F in Table 1) for uncoupled (black lines) and coupled simulations (red lines). **b**, Impact of changes in tropical temperature versus impact of changes in atmospheric CO₂ on tropical land carbon, for the C⁴MIP models (black letters) and three variants of the HadCM3C model¹⁴ (red letters). The horizontal lines represent the new constraint presented in this study.

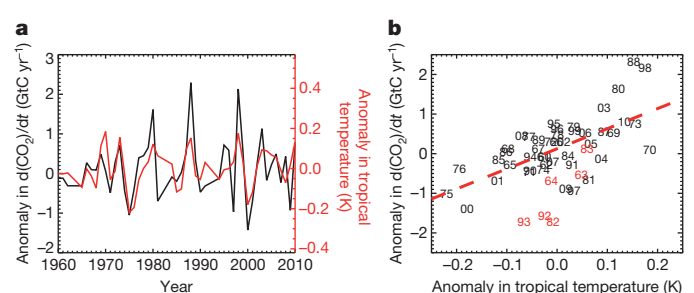


Figure 2 | Observed relationship between variations in the growth rate of atmospheric CO₂ and tropical temperature. **a**, Annual anomalies in CO₂ growth rate (black) and tropical temperature (red) versus year. **b**, Sensitivity of CO₂ growth rate to tropical temperature, with numbers representing the individual years in **a** and the dashed line showing the best-fit straight line, which has a gradient of $5.1 \pm 0.9 \text{ GtC yr}^{-1} \text{ K}^{-1}$. The years in red were not included in this fit because they directly followed major volcanic perturbations to the climate.

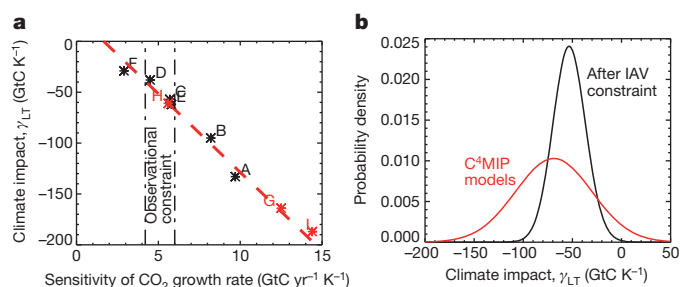


Figure 3 | Emergent constraint on the sensitivity of tropical land carbon to climate change. **a**, Climate sensitivity of tropical land carbon (γ_{LT}) versus the sensitivity of the CO_2 growth rate to tropical temperature, for each of the models shown in Table 1. The dashed line shows the best-fit straight line across the C^4MIP models (black). The red symbols represent a test of this relationship against the three HadCM3C ensemble members. The dot-dash lines indicate the constraint on the observed IAV in the CO_2 growth rate derived from Fig. 2b. **b**, PDF for the climate sensitivity of γ_{LT} . The black line was derived by applying the IAV constraint to the across-model relationship shown in **a**. The red line shows the 'prior' PDF that arises from assuming that all of the C^4MIP models are equally likely to be correct and that they come from a Gaussian distribution.

ensemble members, which were produced by perturbing only parameters in the land carbon-cycle component of the model¹³, span an even larger range ($5.6\text{--}14.4\text{ GtC yr}^{-1}\text{ K}^{-1}$), suggesting that uncertainties in the modelling of the tropical land carbon cycle are critical.

Most importantly, these differing IAV sensitivities are strongly correlated ($r = 0.98$, $P = 0.0005$) with variations in γ_{LT} across C^4MIP models (black labels in Fig. 3a). The dashed red line in Fig. 3a shows the best-fit straight line relating these variables for the six C^4MIP GCMs (although in principle a well-defined nonlinear function would also yield an emergent constraint). The red labels in Fig. 3a show how well this relationship would have predicted the variation in γ_{LT} for the three HadCM3 ensemble members given the IAV sensitivity of each. We note that two of the HadCM3 variants have γ_{LT} values beyond the range of the C^4MIP models, but that the extrapolated straight line is nevertheless able to fit these outliers. The dotted vertical black lines in Fig. 3a show the IAV sensitivity ($\pm 1\text{ s.d.}$), as previously estimated from the contemporary observations, from which we derive tighter bounds on γ_{LT} .

With the model-derived relationship between γ_{LT} and the IAV sensitivity, we can use the observational constraint to estimate a probability density function (PDF) for γ_{LT} (Methods). Figure 3b compares this with the PDF arising from assuming that all C^4MIP models are equally likely to be true and come from an underlying Gaussian distribution (red line). The emergent constraint from the IAV sensitivity of the CO_2 growth rate sharpens the PDF of γ_{LT} and moves its peak to a less negative value (-53 ± 17 as opposed to $-69 \pm 39\text{ GtC K}^{-1}$). The application of the IAV constraint reduces the estimated probability of γ_{LT} values more negative than -100 GtC K^{-1} , typically associated with models that project CO_2 -induced tropical forest dieback, by almost two orders of magnitude from 21% to 0.24%.

The IAV constraint also gives strong confirmation that tropical land carbon is vulnerable to warming caused by non- CO_2 forcing factors¹¹. Remaining uncertainties in tropical land climate-carbon-cycle feedbacks are therefore the magnitude of long-term CO_2 fertilization effects in the tropics, and the extent to which future climate change will be caused by non- CO_2 factors.

METHODS SUMMARY

We used results from six of the eleven models used in C^4MIP^3 . The five excluded models consisted of four Earth-system models of intermediate complexity, which do not typically generate internal variability as required to define the interannual sensitivity of the CO_2 growth rate to tropical temperature anomalies, and one GCM (LLNL), which reported zonal mean land temperatures rather than zonal mean (land and ocean) temperatures. Outputs from the remaining six models were reported as annual means for each 30° latitudinal band (available at https://c4mip.lscce.ipsl.fr/diagnostics_phase2.html).

We combined the outputs from the $30^\circ\text{N--}0^\circ$ and $0^\circ\text{--}30^\circ\text{S}$ bands to define the projected changes for the $30^\circ\text{N--}30^\circ\text{S}$ 'tropical' band.

Models G, H and I in this study, which are used to test the emergent constraint derived from the C^4MIP models, come from a land carbon-cycle ensemble carried out with the HadCM3C model¹⁴. HadCM3C is similar to C^4MIP model A (HadCM3LC) but includes a higher-resolution ocean model ($1.25^\circ \times 1.25^\circ$ rather than $2.5^\circ \times 3.75^\circ$) and interactive atmospheric sulphur-cycle chemistry. Seventeen HadCM3C ensemble members were defined by perturbations to key land surface parameters including leaf nitrogen concentrations and the temperature sensitivities of photosynthesis and soil respiration¹⁴. All ensemble members were driven by the SRES A1B emissions scenarios, including changes in non- CO_2 forcing factors (most notably changes in anthropogenic sulphate aerosols¹⁰). Uncoupled simulations were carried out only for the standard parameter values (HadCM3-st), and the ensemble members leading to the lowest (HadCM3-a) and highest (HadCM3-h) global carbon-cycle feedbacks. We therefore focused on these three variants of HadCM3C in this study.

The analysis of the model outputs and observational data, and the statistical methods employed are outlined in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 31 May; accepted 28 December 2012.

Published online 6 February 2013.

- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
- Malhi, Y. *et al.* Climate change, deforestation, and the fate of the Amazon. *Science* **319**, 169–172 (2008).
- Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C^4MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
- Jupp, T. E. *et al.* Development of probability density functions for future South American rainfall. *New Phytol.* **187**, 682–693 (2010).
- Rammig, A. *et al.* Estimating the risk of Amazonian forest dieback. *New Phytol.* **187**, 694–706 (2010).
- Galbraith, D. *et al.* Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change. *New Phytol.* **187**, 647–665 (2010).
- Hall, A. & Qu, X. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.* **33**, L03502 (2006).
- Bacastow, R. Modulation of atmospheric carbon dioxide by the Southern Oscillation. *Nature* **261**, 116–118 (1976).
- Cox, P. M. *et al.* Amazon dieback under climate-carbon cycle projections for the 21st century. *Theor. Appl. Climatol.* **78**, 137–156 (2004).
- Cox, P. M. *et al.* Increasing risk of Amazonian drought due to decreasing aerosol pollution. *Nature* **453**, 212–215 (2008).
- Huntingford, C. *et al.* Highly contrasting effects of different climate forcing agents on ecosystem services. *Phil. Trans. R. Soc. A* **369**, 2026–2037 (2011).
- Nakicenovic, N. *et al.* Emissions Scenarios: Summary for Policymakers. Spec. Report (Intergovernmental Panel on Climate Change, 2000).
- Friedlingstein, P., Dufresne, J.-L., Cox, P. M. & Rayner, P. How positive is the feedback between climate change and the carbon cycle? *Tellus* **55B**, 692–700 (2003).
- Booth, B. B. *et al.* High sensitivity of future global warming to land carbon cycle processes. *Environ. Res. Lett.* **7**, 024002 (2012).
- Moss, R. H. *et al.* The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756 (2010).
- Hungate, B. A. *et al.* Nitrogen and climate change. *Science* **302**, 1512–1513 (2003).
- Zaehle, S., Friedlingstein, P. & Friend, A. D. Terrestrial nitrogen feedbacks may accelerate future climate change. *Geophys. Res. Lett.* **37**, L01401 (2010).
- Baker, T. R. *et al.* Increasing biomass in Amazonian forest plots. *Phil. Trans. R. Soc. Lond. B* **359**, 353–365 (2004).
- Lewis, S. L. *et al.* Increasing carbon storage in intact African tropical forests. *Nature* **457**, 1003–1006 (2009).
- Marengo, J. A. *et al.* The drought of Amazonia in 2005. *J. Clim.* **21**, 495–516 (2008).
- Marengo, J. A. *et al.* The drought of 2010 in the context of historical droughts in the Amazon region. *Geophys. Res. Lett.* **38**, L12703 (2011).
- Phillips, O. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344–1347 (2009).
- Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
- Jones, C. D. & Cox, P. M. On the significance of atmospheric CO_2 growth rate anomalies in 2002–2003. *Geophys. Res. Lett.* **32**, L14816 (2005).
- Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 499–587 (Cambridge Univ. Press, 2007).
- Masarie, K. A. & Tans, P. P. Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record. *J. Geophys. Res.* **100**, 11593–11610 (1995).
- Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2500. *Clim. Change* **109**, 213–241 (2011).

28. Smith, T. M. *et al.* Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.* **21**, 2283–2296 (2008).
29. Mercado, L. M. *et al.* Impact of changes in diffuse radiation on the global land carbon sink. *Nature* **458**, 1014–1017 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge funding from the NERC NCEO programme (P.M.C. and C.M.L.); the EU Greencycles II project (P.M.C. and P.F.); the EU FP7 'CARBONES' project (D.P. and C.D.J.); the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101) (D.P., B.B.B. and C.D.J.); the CEH Science Budget (C.H.) and the Newton Institute programme on 'Mathematical and Statistical Approaches to Climate Modelling and Prediction', during which this research was first formulated (P.M.C., B.B.B. and C.H.). We also acknowledge the modelling groups that provided results to C⁴MIP.

Author Contributions P.M.C. led the study and drafted the manuscript. D.P. assisted with the statistical analysis, especially the estimation of the observationally constrained PDF in Fig. 3b. P.F. provided data and guidance on the C⁴MIP model ensemble, and B.B.B. did likewise for the HadCM3 carbon-cycle ensemble. C.H. processed observational climate data sets to produce time series of tropical mean temperature anomalies. P.M.C., C.D.J., P.F. and C.H. have had discussions over many years concerning the relationship between the interannual variability and the long-term sensitivity of the land carbon cycle to climate change. C.M.L. provided invaluable insights into the interpretation of the regression line in Fig. 3a. All co-authors commented on and provided edits to the original manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.M.C. (p.m.cox@exeter.ac.uk).

METHODS

Choice of models and variables. To make use of the observed interannual variation in atmospheric CO₂ as a constraint, we need climate–carbon-cycle simulations that model CO₂ as a ‘free’, fully prognostic variable. We therefore make use of the C⁴MIP simulations³, which used prescribed SRES A2 CO₂ emissions but calculated the global mean atmospheric CO₂ concentration interactively. We have augmented the C⁴MIP results with free CO₂ runs from a carbon-cycle parameter ensemble carried out with HadCM3 (ref. 14). These HadCM3 runs allow the emergent constraint derived from the C⁴MIP models to be tested over a wide range of possible future carbon losses from tropical land.

To derive an emergent constraint, it is of paramount importance that equivalent variables are compared from the models and observations. Therefore, because the C⁴MIP models reported global mean atmospheric CO₂, and mean land-plus-ocean near-surface temperatures, we compute the same diagnostics from the observational data sets (see below).

Diagnosis of γ_{LT} . The sensitivity of tropical land carbon storage to temperature, γ_{LT} , is calculated as in previous studies^{3,13}. First, the sensitivity of tropical land carbon storage to direct CO₂ effects, as given by the parameter β_{LT} , is diagnosed from the uncoupled simulation for each model, $\beta_{LT} = \Delta C_{LT}^u / \Delta C_a^u$, where $\Delta C_{LT}^u = C_{LT}^u(t_1) - C_{LT}^u(t_0)$ is the change in tropical land carbon storage (in GtC) and $\Delta C_a^u = C_a^u(t_1) - C_a^u(t_0)$ is the change in global atmospheric CO₂ concentration (in p.p.m.v.), in both cases between time t_0 and time t_1 for the uncoupled simulation.

This value of β_{LT} is then used to isolate γ_{LT} from the coupled simulation of each model, using the equation

$$\gamma_{LT} = \frac{\Delta C_{LT}^c - \beta_{LT} \Delta C_a^c}{\Delta T_T^c}$$

where $\Delta C_{LT}^c = C_{LT}^c(t_1) - C_{LT}^c(t_0)$ is the change in tropical land carbon storage (in GtC), $\Delta C_a^c = C_a^c(t_1) - C_a^c(t_0)$ is the change in global atmospheric CO₂ concentration (in p.p.m.v.) and $\Delta T_T^c = T_T^c(t_1) - T_T^c(t_0)$ is the change in mean tropical (30° N–30° S) temperature (in K), in all cases between time t_0 and time t_1 for the coupled simulation.

We define the changes relative to 1960 in all cases (that is, $t_0 = 1960$), and use the longest possible common simulation periods over which to diagnose β_{LT} and γ_{LT} for the C⁴MIP models ($t_1 = 2099$) and the HadCM3C ensemble members ($t_1 = 2080$), respectively.

Sensitivity of CO₂ growth-rate anomaly to tropical temperature anomaly. The sensitivity of the atmospheric CO₂ growth rate to tropical temperature is calculated over the period 1960–2010 inclusive, for the observations and all models. However, for the observational data, and the HadCM3C simulations, which included volcanoes, we exclude the years 1963, 1964, 1982, 1983, 1991 and 1992, which were heavily influenced by the volcanic eruptions²⁴ of Mount Agung (in 1963), El Chichon (in 1982) and Mount Pinatubo (in 1991). There are two reasons for removing volcanoes. First, not all the models in our ensembles include the climatic effects of volcanic eruptions. Second, volcanoes are believed to affect the land carbon sink through the effects of diffuse radiation fertilization²⁹, but these effects are not included in the generation of models considered here. We therefore removed ‘volcano years’ from the observations to maximize consistency between models and observations.

For comparability with the outputs available from the C⁴MIP models, we also use the global CO₂ concentration and the mean tropical (30° N–30° S) temperature, including both land and ocean points.

As in previous studies²⁴, the annual CO₂ growth rate for the n th year, $dC_a/dt(t_n)$, is defined as the difference between the annual mean CO₂ concentrations for the n th and $(n-1)$ th years: $dC_a/dt(t_n) = C_a(t_n) - C_a(t_{n-1})$. The CO₂ growth rate is therefore centred in time at the beginning of year n . To align the tropical temperature anomalies, we take the associated tropical mean temperature, $T_T(t_n)$, to be the mean of the annual mean tropical temperatures for years n and $n-1$:

$$\bar{T}_T(t_n) = \frac{T_T(t_n) + T_T(t_{n-1})}{2}$$

For all model and observational time series, the annual CO₂ growth rate, dC_a/dt , and the associated mean tropical temperature, \bar{T}_T , were detrended using an 11-yr running mean, with the residuals defining the annual anomalies (Supplementary Fig. 2). In each case, a least-squares linear regression was found between these anomalies in the CO₂ growth rate and the anomalies in the tropical temperature, with the gradient of the best fit defining the IAV sensitivity (see below).

The IAV sensitivity was calculated for a range of data sets of tropical temperature and atmospheric CO₂ (see below), to explore the uncertainty in the estimate of the IAV sensitivity arising from uncertainties in the observational data. These different estimates are listed in Supplementary Table 1.

To isolate the separate contributions of the tropical net primary productivity and soil respiration, similar regressions against tropical temperature anomalies were carried out separately for each of these fluxes as diagnosed from the C⁴MIP models (Supplementary Fig. 1). This showed that the IAV sensitivity across the model ensemble is correlated with the response of tropical soil respiration (Supplementary Fig. 1b), rather than net primary productivity (Supplementary Fig. 1c). By contrast, the wide range of longer-term projections of changes in land carbon storage is known to be in part due to the different responses of net primary productivity to climate change³.

Observational data. Observed annual global CO₂ concentration²⁶ for 1980 to 2010 was downloaded from the NOAA website (http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html#global_data). Because this data set covers only the period from 1980, global CO₂ concentrations for 1960–1979 were taken from the historical data sets derived for use with the RCP scenarios²⁷ (<http://www.pik-potsdam.de/~mmalte/rcps/index.htm#>).

Tropical (30° N–30° S) annual mean temperatures were calculated from NCDC data²⁸ (<http://www.ncdc.noaa.gov/ghcnm/maps.php>), from the CRU/Met Office HadCRU3 data set (<http://www.metoffice.gov.uk/hadobs/hadcrut3/>) and from the GISS data set (<http://data.giss.nasa.gov/gistemp/>).

Least-squares linear regression. Least-squares linear regressions were calculated using well-established formulae (see, for example, <http://mathworld.wolfram.com/LeastSquaresFitting.html>). The linear regression, f_n , between a time series given by y_n and a time series given by x_n is defined by a gradient, b , and intercept, a : $f_n = a + bx_n$. Minimizing the least squares error for y_n involves minimizing

$$s^2 = \frac{1}{N-2} \sum_{n=1}^N \{y_n - f_n\}^2$$

where N is the number of data points in each time series. In this case, the best-fit gradient is given by $\hat{b} = \sigma_{xy} / \sigma_x^2$. Here $\sigma_x^2 = \sum_{n=1}^N \{x_n - \bar{x}\}^2 / N$ is the variance of x_n and $\sigma_{xy} = \sum_{n=1}^N \{x_n - \bar{x}\} \{y_n - \bar{y}\} / N$ is the covariance of the x_n and y_n time series, which have means of \bar{x} and \bar{y} , respectively. The standard error of b is given by $\sigma_b = s / \sigma_x \sqrt{N}$, which defines a Gaussian probability density for b :

$$P(b) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{-\frac{(b-\hat{b})^2}{2\sigma_b^2}\right\}$$

The ‘prediction error’ of the regression is the following function of x :

$$\sigma_f(x) = s \sqrt{1 + \frac{1}{N} + \frac{(x-\bar{x})^2}{N\sigma_x^2}}$$

This expression defines contours of equal probability density around the best-fit linear regression, which represent the probability density of y given x :

$$P\{y|x\} = \frac{1}{\sqrt{2\pi}\sigma_f} \exp\left\{-\frac{(y-f(x))^2}{2\sigma_f^2}\right\}$$

where $\sigma_f = \sigma_f(x)$, as above.

Calculation of the PDF for γ_{LT} . The emergent constraint derived in this study is a linear regression across the C⁴MIP GCMs between the temperature sensitivity of land carbon storage in the tropics, γ_{LT} , and the sensitivity of the annual growth rate in atmospheric CO₂ to the annual tropical temperature anomaly, which we label here as γ_{CO_2} . In the context of the least-squares linear regression presented above, γ_{LT} is equivalent to y and γ_{CO_2} is equivalent to x .

The linear regression therefore provides an equation for the probability of γ_{LT} given γ_{CO_2} (that is, the equation for $P\{y|x\}$ above). Supplementary Fig. 3 shows the best-fit straight line (thick dashed red line) and the plus and minus σ_f prediction error contours (as thin dashed red lines) on the same scales as in Fig. 3a.

In addition, the linear regression between the observed annual anomalies in the atmospheric CO₂ growth rate^{25,26} and the tropical mean temperature²⁷ provides an observation-based PDF for γ_{CO_2} (via the equation for $P(b)$ above). The best-fit γ_{CO_2} from these observations is shown by the thick dashed vertical line in Supplementary Fig. 3, and the uncertainty in this fit is shown by the thin dashed vertical lines representing plus and minus 1 s.e. about the best-fit value.

Given these two PDFs, $P\{\gamma_{LT}|\gamma_{CO_2}\}$ and $P(\gamma_{CO_2})$, the PDF for γ_{LT} is

$$P(\gamma_{LT}) = \int_{-\infty}^{\infty} P\{\gamma_{LT}|\gamma_{CO_2}\} P(\gamma_{CO_2}) d\gamma_{CO_2}$$

The integrand, $P\{\gamma_{LT}|\gamma_{CO_2}\} P(\gamma_{CO_2})$, is shown by the continuous black contours in Supplementary Fig. 3, and the integral is the basis for the black PDF for γ_{LT} shown in Fig. 3b.

Behavioural and genetic analyses of *Nasonia* shed light on the evolution of sex pheromones

Oliver Niehuis^{1,2}, Jan Buellesbach^{2,3,4}, Joshua D. Gibson², Daniela Pothmann⁵, Christian Hanner³, Navdeep S. Mutti^{2,6}, Andrea K. Judson², Jürgen Gadau², Joachim Ruther⁵ & Thomas Schmitt^{3,4,7}

Sex pheromones play a pivotal role in the communication of many sexually reproducing organisms¹. Accordingly, speciation is often accompanied by pheromone diversification enabling proper mate finding and recognition². Current theory implies that chemical signals are under stabilizing selection by the receivers who thereby maintain the integrity of the signals³. How the tremendous diversity of sex pheromones seen today evolved is poorly understood^{4,5}. Here we unravel the genetics of a newly evolved pheromone phenotype in wasps and present results from behavioural experiments indicating how the evolution of a new pheromone component occurred in an established sender–receiver system. We show that male *Nasonia vitripennis* evolved an additional pheromone compound differing only in its stereochemistry from a pre-existing one. Comparative behavioural studies show that conspecific females responded neutrally to the new pheromone phenotype when it evolved. Genetic mapping and gene knockdown show that a cluster of three closely linked genes accounts for the ability to produce this new pheromone phenotype. Our data suggest that new pheromone compounds can persist in a sender's population, without being selected against by the receiver and without the receiver having a pre-existing preference for the new pheromone phenotype, by initially remaining unperceived. Our results thus contribute valuable new insights into the evolutionary mechanisms underlying the diversification of sex pheromones. Furthermore, they indicate that the genetic basis of new pheromone compounds can be simple, allowing them to persist long enough in a population for receivers to evolve chemosensory adaptations for their exploitation.

Sexually reproducing organisms depend on their ability to localize and recognize conspecific mates reliably¹. Many insects, for example, make use of sex pheromones to attract mates. The signal integrity of sex pheromones has been postulated to be maintained by pheromone receivers selecting against modifications of a given pheromone composition. It is unclear, however, how sex pheromones diversify among species and how new pheromone phenotypes evolve if stabilizing selection is acting against pheromone modifications^{4,5}. Studies on *Drosophila* and moths have provided important insights into the evolution of quantitative pheromone changes (shifts in the relative abundance of individual pheromone components). Notably, they have highlighted the importance of pre-existing pheromone preferences and biosynthetic enzymes in the evolution of pheromone diversity^{2,6–12}. To shed light on the evolution of qualitative pheromone alterations, we studied sex pheromones in the wasp genus *Nasonia*, an emerging model system for investigating the genetics of speciation and complex traits^{13,14}.

Male *Nasonia vitripennis* attract conspecific virgin females by releasing a sex pheromone consisting of 4(R),5(S)-5-hydroxy-4-decanolide (referred to here as RS), 4(R),5(R)-5-hydroxy-4-decanolide

(referred to here as RR) and 4-methylquinazoline (referred to here as MQ)^{15–18} (Fig. 1a, b, d). We compared this pheromone with those of all other known *Nasonia* species and with that of the closely related wasp *Trichomalopsis sarcophagae* and found *N. vitripennis* to be the only species whose males biosynthesize RR (Fig. 1b and Supplementary Fig. 1). The pheromone of all other *Nasonia* species and of *T. sarcophagae* consists only of RS and MQ. Therefore the ancestral sex pheromone of *Nasonia* males probably consisted solely of RS and MQ, and the additional occurrence of RR is probably a derived state that evolved in the *N. vitripennis* lineage (Fig. 1c).

To assess the contribution of RR as an active component of the *N. vitripennis* sex pheromone, we exposed virgin females of *N. vitripennis* and *Nasonia giraulti* (representative of the sister lineage of *N. vitripennis* occurring in sympatry with *N. vitripennis*^{13,14}; Fig. 1c) in a two-choice olfactometer to natural pheromone extracts of males of these two species and to synthetic pheromone components. Our experiments showed that females are attracted by natural pheromone extracts of both con- and heterospecific males and do not differentiate between natural extracts and synthetic pheromone blends (Fig. 1e, f and Supplementary Tables 1 and 2). Females of both species were also attracted by 100-ng doses of synthetic RS (Fig. 1e, f). However, females of neither species responded to 100-ng doses of synthetic RR alone (Fig. 1e, f). We subsequently manipulated sex pheromone extracts of *N. giraulti* males by adding synthetic RR at naturally occurring dosages, thereby creating a pheromone phenotype almost identical to that of *N. vitripennis* males. When given the choice between manipulated and genuine pheromone extract, *N. giraulti* females did not discriminate between them (Fig. 1g and Supplementary Fig. 2). *N. vitripennis* females, however, clearly preferred the manipulated *N. giraulti* extract containing synthetic RR. Hence, RR influenced the behaviour of *N. vitripennis* females only in the presence of the other pheromone components. In fact, their response to the *N. giraulti*-manipulated pheromone extract did not differ significantly from their response to *N. vitripennis* pheromone extract (Fig. 1g). This notable result opens up the possibility that *N. vitripennis* females did not respond to RR when it first occurred as a new component of their males' sex pheromone—a behaviour still found today in the sister lineage of *N. vitripennis*. Thus, instead of having had a predisposition to prefer the new pheromone blend, even before it existed, *N. vitripennis* females may have evolved this preference later, fostered by the co-occurrence of RR and RS.

To determine the genetic basis of RR biosynthesis, we studied recombinant F₂ hybrid males of *N. vitripennis* and *N. giraulti*. We found the RR/RS ratio in the sex pheromone of the hybrids to vary across a wide range and exploited this variation in a quantitative trait locus analysis, studying molecular markers distributed across all five chromosomes of *Nasonia*. The analysis indicated two quantitative trait loci (QTL) at a genome-wide significance level of 0.01: one on

¹Centre for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany. ²School of Life Sciences, Arizona State University, Tempe, Arizona 85287, USA.

³Evolutionary Biology and Animal Ecology, University of Freiburg, 79104 Freiburg, Germany. ⁴Spemann Graduate School of Biology and Medicine, University of Freiburg, 79104 Freiburg, Germany.

⁵Department of Zoology, University of Regensburg, 93053 Regensburg, Germany. ⁶DuPont Experimental Station, DuPont-Pioneer Agricultural Biotechnology, Wilmington, Delaware 19880-0353, USA.

⁷Ecological Network Group, Technical University of Darmstadt, 64287 Darmstadt, Germany.

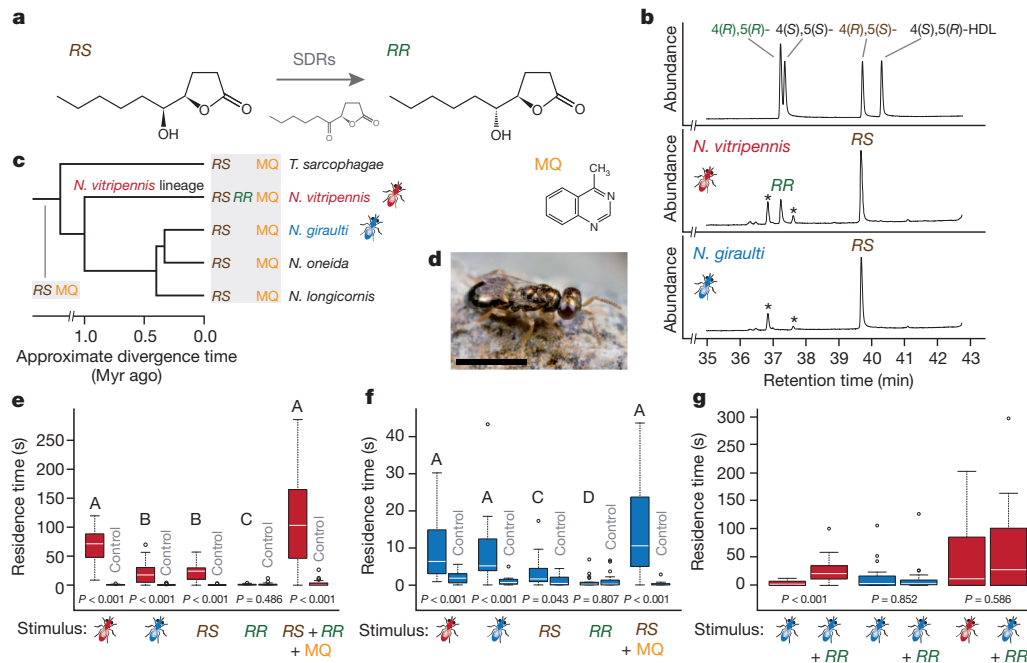


Figure 1 | Evolution of sex pheromone diversity and behavioural response in *Nasonia* parasitoid wasps. **a**, Chemical structures of *Nasonia* male-specific sex pheromone compounds. In grey: proposed biosynthesis of 4(R),5(R)-5-hydroxy-4-decanolide (RR) by oxidation and successive reduction of 4(R),5(S)-5-hydroxy-4-decanolide (RS) catalysed by short-chain dehydrogenases/reductases (SDRs). **b**, Enantioselective gas chromatography–mass spectrometry of synthetic samples of all four stereoisomers of 5-hydroxy-4-decanolide (HDL) (top) and sex pheromone extracts of *N. vitripennis* (middle) and *N. giraulti* males (bottom). Asterisks indicate contaminants. **c**, Phylogenetic relationships of *Nasonia* wasps and evolutionary history of male-specific sex pheromone components. **d**, Adult *N. vitripennis* male (scale bar, 1 mm). **e**, **f**, Box-and-whisker plots showing median (horizontal line), interquartile range (box), maximum/minimum range (whiskers) and outliers ($>1.5\times$ upper quartile) of the residence time spent by virgin *N. vitripennis*

chromosome 1 (log of odds (lod) score = 42.7; $P < 0.001$) and the other on chromosome 4 (lod score = 5.2; $P < 0.001$) (Fig. 2a and Supplementary Fig. 3). The predicted QTL on chromosome 1 explained 81.5% of the total phenotypic variance and was indicative of whether or not the *N. vitripennis* pheromone phenotype was observed, whereas the QTL on chromosome 4 explained only 3.9% of the variance.

(e) and *N. giraulti* (f) females in the two odour fields of a static two-choice olfactometer when given the choice between a solvent control, pheromone extracts of *N. vitripennis* males (red wasp) and *N. giraulti* males (blue wasp) and enantiopure synthetic pheromone components (RS, RR and MQ) (Wilcoxon matched-pairs test; $n = 20$ per test). Differences between treatments (residence time in the test field) were assessed with the Kruskal–Wallis test followed by pairwise comparisons with Holm-corrected Mann–Whitney U tests (different uppercase letters indicate statistical differences at $P < 0.05$). **g**, Response of *N. vitripennis* (red boxes) and *N. giraulti* (blue boxes) females to sex pheromone extract (representing 100 ng of RS) of male *N. vitripennis* and male *N. giraulti*, and to the same amount of male *N. giraulti* sex pheromone extract to which 60 ng of synthetic RR had been added (blue wasp + RR). Statistical analysis and illustration as in e and f.

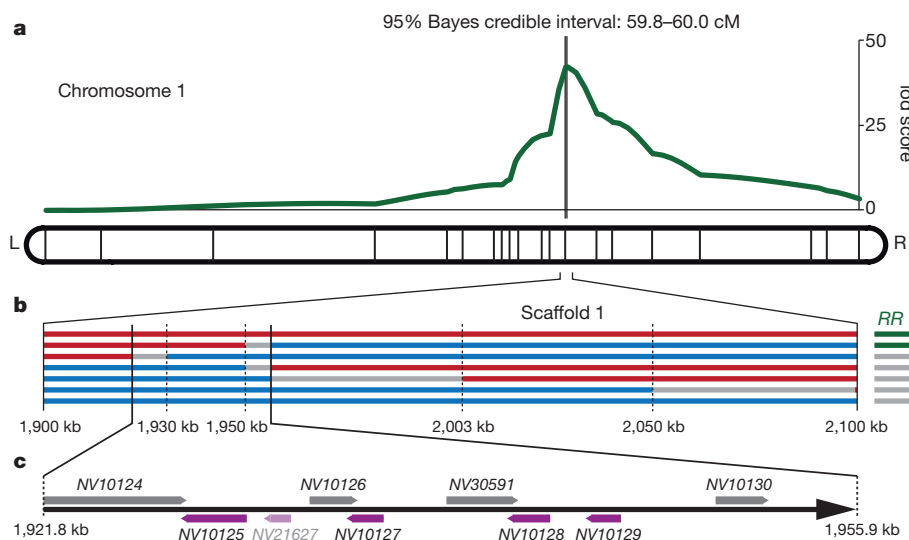


Figure 2 | Genetics of sex pheromone differences between *N. vitripennis* and *N. giraulti* males.

a, Mapping of QTL that explain the RR/RS ratio in the sex pheromones of *N. vitripennis* \times *N. giraulti* F₂ hybrid males. The curve specifies the lod score along chromosome 1 for the presence of a genetic factor responsible for the RR/RS ratio of hybrid offspring. **b**, Fine mapping of the genetic factor on genome sequence scaffold 1 of chromosome 1 that controls for the presence of RR in the males' sex pheromone. Hybrids with the *N. vitripennis* genotype (red) between nucleotides 1,921.8 kb and 1,955.9 kb on scaffold 1 have RR present in the sex pheromone, those with the *N. giraulti* (blue) genotype in this region lack RR. In grey are regions with unknown genotype. **c**, Predicted genes in the candidate region of genome sequence scaffold 1 of chromosome 1. In violet are predicted SDRs (the predicted gene NV21627 is not supported by extrinsic evidence, for example, transcript data).

N. vitripennis genotype and the presence of *RR* held across all recombinant hybrids. This allowed us to narrow down the location of the QTL to a 34.1-kb region (Fig. 2b).

The official gene set for *N. vitripennis*¹⁹ predicted nine genes in the candidate region (Fig. 2c). To learn more about their potential functions, we searched their amino acid sequences against the Pfam protein family database²¹. Intriguingly, five predicted genes showed high similarity to short-chain dehydrogenases/reductases (SDRs), especially 15-hydroxyprostaglandin dehydrogenases: *NV10125*, *NV10127*, *NV10128*, *NV10129* and *NV21627*. SDRs are oxidoreductases with epimerase and/or racemase activity that catalyse the oxidation of hydroxyl groups to the respective carbonyls and vice-versa^{22,23}. Given that *RS* and *RR* differ only in the chirality of their hydroxyl group, we considered these genes promising candidates for mediating a conversion of *RS* to *RR*.

Gene expression data provided evidence for the expression of only four of the five candidate genes in adult *N. vitripennis* males: *NV10125*, *NV10127*, *NV10128* and *NV10129*. Of these, *NV10127*, *NV10128* and *NV10129* showed an astonishingly high similarity at the nucleotide sequence level (87.4–98.7% identity). Phylogenetic analysis subsequently indicated that this tandem array of genes—all of which are present in *N. giraulti* and similarly expressed in *N. vitripennis* and *N. giraulti* (Supplementary Fig. 4 and Supplementary Table 3)—was subject to gene conversion (the information of one gene was partially or completely replaced by that of another) in the *N. vitripennis* lineage (Fig. 3a and Supplementary Fig. 5), which is a possible explanation for changes in gene function linked to the evolution of a novel pheromone compound. We therefore decided to test the involvement of *NV10127*, *NV10128* and *NV10129* in the biosynthesis of *RR* by means of double-stranded RNA (dsRNA)-mediated gene knockdown experiments. We injected dsRNA of *NV10127* (treatment A) and *NV10128* (treatment B) as well as a mixture of both (treatment AB) into male pupae of *N. vitripennis*. We did not synthesize or inject dsRNA of *NV10129* as the coding sequence of the gene was almost identical to that of *NV10128* in the 539-base pair target region for the dsRNA of *NV10128*. Gene expression analysis in the treated males subsequently confirmed significant (71–93%) knockdown of *NV10127*, *NV10128* and *NV10129* relative to controls in which we injected dsRNA derived from green fluorescent protein (GFP)-encoding sequence (treatment C) (Fig. 3b–d). None of the treatments significantly altered expression of *NV10125*, the only remaining candidate gene expressed in adult males ($P = 0.296$; Fig. 3e). Intriguingly, treatments A, B and AB caused a significant shift in the *RR/RS* ratio towards *RS* in the pheromone blend of adult males compared to controls (Fig. 3f). This establishes that at least one of the three predicted SDR-encoding genes is required for the ability to produce the *N. vitripennis* pheromone phenotype. We found no statistically significant differences in the knockdown and *RR/RS* ratio change between the different treatments that could have given further insights into the importance of each of the three putative SDRs ($P \geq 0.275$).

This is the first study, to our knowledge, on the genetics of pheromone evolution in Hymenoptera (wasps, ants and bees) and the first characterization of genes that cause a qualitative change in an insect pheromone by triggering the stereochemical inversion of a pre-existing signalling molecule. The results shed new light on the biosynthesis and evolution of chemical signals in sender–receiver systems. The high similarity of the identified putative SDRs with 15-hydroxyprostaglandin dehydrogenases (oxidoreductases with carbonyl reductase activity targeting the 15(*S*)-hydroxy group of prostaglandins²⁴) and the structural resemblance between 15-hydroxyprostaglandins and 5-hydroxy-4-decanolide (a five-membered oxygenated ring with a hydroxylated side chain) suggest that the putative SDRs catalyse the direct conversion of *RS* to *RR*. Specifically, we assume that the hydroxyl group at carbon atom five of *RS* is oxidized by one of the three putative SDRs and that the resulting ketone is subsequently reduced with inversion of stereochemistry by either the same or one of the other two

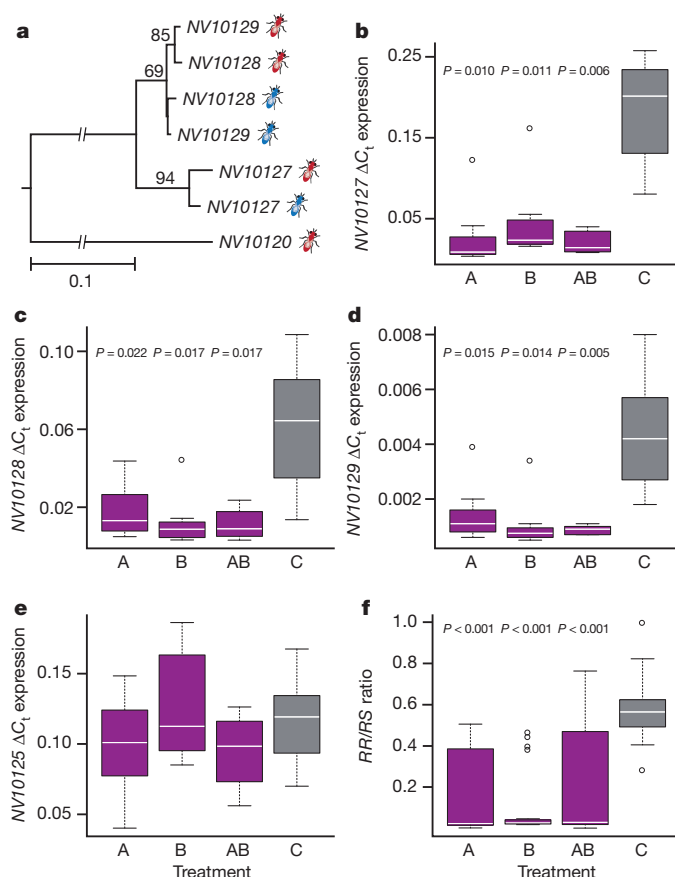


Figure 3 | Phylogeny and dsRNA-mediated knockdown of candidate genes. **a**, Phylogenetic relationships of the predicted SDR-encoding genes *NV10127*, *NV10128* and *NV10129* of *N. vitripennis* (red wasps) and *N. giraulti* (blue wasps) inferred from coding sequence information. *NV10120* of *N. vitripennis* was used for outgroup comparison. Numbers above branches indicate statistical bootstrap support. **b**, Box-and-whisker plot showing median (horizontal line), interquartile range (box), maximum/minimum range (whiskers) and outliers ($>1.5 \times$ upper quartile) of the expression of *NV10127* in male *N. vitripennis* after injecting dsRNA of *NV10127* (treatment A), *NV10128* (treatment B), equimolar mixture of *NV10127* and *NV10128* (treatment AB), or dsRNA derived from GFP sequence (treatment C) (Kruskal–Wallis test followed by pairwise comparisons with Holm-corrected Mann–Whitney *U* tests; $n = 8$ per treatment). **c–e** All samples treated, statistically analysed and illustrated as in **b**. **c**, Expression of *NV10128* in male *N. vitripennis*. **d**, Expression of *NV10129* in male *N. vitripennis*. **e**, Expression of *NV10125* in male *N. vitripennis*. ($P = 0.296$; Kruskal–Wallis test; $n = 8$ per treatment.) **f**, The *RR/RS* ratio in sex pheromone extracts of *N. vitripennis* males treated and illustrated as in **b** (Kruskal–Wallis test followed by pairwise comparisons with Holm-corrected Mann–Whitney *U* tests; $n = 24$ –52 per treatment.)

putative SDRs, leading to *RR* as a new pheromone component (Fig. 1a). Our data thus provide support for the idea that some new pheromone compounds arise as modifications of existing signalling molecules⁴, and show that the genetic basis of these modifications can be relatively simple. They further indicate that sex pheromone receivers do not automatically respond to new pheromone compounds within an existing pheromone blend. Thus, a new pheromone component may remain initially uncoupled from selective constraints exerted by the receiver. This and a simple genetic basis underlying the biosynthesis of a new compound would give a receiver time to associate the new compound with conspecific mates and to recognize it as part of the species-specific chemical signal, thus illustrating how new pheromone compounds can evolve within established sender–receiver systems.

METHODS SUMMARY

We screened several strains of each *Nasonia* species to infer the composition of the male-specific sex pheromone using coupled gas chromatography–mass

spectrometry. The stereochemistry of 5-hydroxy-4-decanolide isomers was established by enantioselective gas chromatography¹⁵. We tested the response of virgin *N. vitripennis* and *N. giraulti* females to naturally occurring doses of purified pheromone extracts of either species and to synthetic pheromone blends containing RS, RR and MQ using a two-choice olfactometer¹⁸. Virgin *N. vitripennis* and *N. giraulti* females were also given the choice between natural *N. giraulti* pheromone extract and manipulated *N. giraulti* pheromone extract that contained synthetic RR. Finally, we tested the response of virgin *N. vitripennis* females when exposed simultaneously to experimentally designed (natural *N. giraulti* extract plus synthetic RR) and genuine *N. vitripennis* pheromone. Bioassay data (residence times in test and control area) were statistically analysed with the Wilcoxon matched-pairs test ($n = 20$ per test), and differences between treatments were assessed with the Kruskal–Wallis test followed by pairwise comparisons with Holm-corrected Mann–Whitney *U* tests. Genetic cross experiments were conducted as described previously²⁵. The QTL analysis was done with R/qtl²⁶, applying a multiple-QTL model and genotyping 71 molecular markers listed in ref. 27 and in Supplementary Table 4 in 106 F₂ hybrid males. We followed the protocol described previously²⁸ for gene knockdown and measured gene expression by quantitative reverse transcriptase PCR (qRT–PCR), applying the oligonucleotide primers listed in Supplementary Tables 5 and 6. Gene expression differences were statistically analysed with the Kruskal–Wallis test followed by pairwise comparisons with Holm-corrected Mann–Whitney *U* tests (for qRT–PCR experiments $n = 8–10$ per locus, treatment and sex; for RR/RS ratio experiments $n = 24–52$ per treatment). Phylogenetic analyses of DNA sequences were executed with MEGA5 (ref. 29) and applying the maximum likelihood optimality criterion. Node support was statistically assessed from 1,000 bootstrap replicates.

Received 30 August; accepted 12 December 2012.

Published online 13 February 2013.

- Wyatt, T. D. *Pheromones and Animal Behaviour* Ch. 3 37–73 (Cambridge Univ. Press, 2003).
- Smadja, C. & Butlin, R. K. On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity* **102**, 77–97 (2009).
- Butlin, R. K. & Trickett, A. J. in *Insect Pheromone Research: New Directions* (eds Cardé, R. T. & Minks, A. K.) 548–562 (Chapman & Hall, 1997).
- Symonds, M. R. E. & Elgar, M. A. The evolution of pheromone diversity. *Trends Ecol. Evol.* **23**, 220–228 (2008).
- Steiger, S., Schmitt, T. & Schaefer, H. M. The origin and dynamic evolution of chemical information transfer. *Proc. R. Soc. B* **278**, 970–979 (2011).
- Roelofs, W. L. & Rooney, A. P. Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *Proc. Natl Acad. Sci. USA* **100**, 9179–9184 (2003).
- Ferveur, J.-F. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behav. Genet.* **35**, 279–295 (2005).
- Xue, B., Rooney, A. P., Kajikawa, M., Okada, N. & Roelofs, W. L. Novel sex pheromone desaturases in the genomes of corn borers generated through gene duplication and retroposition fusion. *Proc. Natl Acad. Sci. USA* **104**, 4467–4472 (2007).
- Shirangi, T. R., Dufour, H. D., Williams, T. M. & Carroll, S. B. Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. *PLoS Biol.* **7**, e1000168 (2009).
- Lassance, J.-M. *et al.* Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* **466**, 486–489 (2010).
- Liénard, M. A., Hagström, A. K., Lassance, J.-M. & Löfstedt, C. Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene. *Proc. Natl Acad. Sci. USA* **107**, 10955–10960 (2010).
- Albre, J. *et al.* Sex pheromone evolution is associated with differential regulation of the same desaturase gene in two genera of leafroller moths. *PLoS Genet.* **8**, e1002489 (2012).
- Liebig, D. W. The parasitoid wasp *Nasonia*: an emerging model system with haploid male genetics. *Cold Spring Harb. Protocols* **2009**, pdb.emo134 (2009).
- Werren, J. H. *et al.* Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**, 343–348 (2010).
- Ruther, J., Stahl, L. M., Steiner, S., Garbe, L. A. & Tolasch, T. A male sex pheromone in a parasitic wasp and control of the behavioral response by the female's mating status. *J. Exp. Biol.* **210**, 2163–2169 (2007).
- Ruther, J., Steiner, S. & Garbe, L.-A. 4-Methylquinazoline is a minor component of the male sex pheromone in *Nasonia vitripennis*. *J. Chem. Ecol.* **34**, 99–102 (2008).
- Abdel-latif, M., Garbe, L. A., Koch, M. & Ruther, J. An epoxide hydrolase involved in the biosynthesis of an insect sex attractant and its use to localize the production site. *Proc. Natl Acad. Sci. USA* **105**, 8914–8919 (2008).
- Steiner, S. & Ruther, J. Mechanism and behavioral context of male sex pheromone release in *Nasonia vitripennis*. *J. Chem. Ecol.* **35**, 416–421 (2009).
- Munoz-Torres, M. C. *et al.* Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. *Nucleic Acids Res.* **39**, D658–D662 (2011).
- Niehuus, O. *et al.* Recombination and its impact on the genome of the haplodiploid parasitoid wasp *Nasonia*. *PLoS ONE* **5**, e8597 (2010).
- Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301 (2012).
- Tanner, M. E. Understanding nature's strategies for enzyme-catalyzed racemization and epimerization. *Acc. Chem. Res.* **35**, 237–246 (2002).
- Kavanagh, K. L., Jönvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **65**, 3895–3906 (2008).
- Cho, H., Oliveira, M. A. & Tai, H.-H. Critical residues for the coenzyme specificity of NAD⁺-dependent 15-hydroxyprostaglandin dehydrogenase. *Arch. Biochem. Biophys.* **419**, 139–146 (2003).
- Niehuus, O., Judson, A. K. & Gadau, J. Cytonuclear genic incompatibilities cause increased mortality in male F₂ hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics* **178**, 413–426 (2008).
- Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
- Niehuus, O., Büllesbach, J., Judson, A. K., Schmitt, T. & Gadau, J. Genetics of cuticular hydrocarbon differences between males of the parasitoid wasps *Nasonia giraulti* and *Nasonia vitripennis*. *Heredity* **107**, 61–70 (2011).
- Lynch, J. A. & Desplan, C. A method for parental RNA interference in the wasp *Nasonia vitripennis*. *Nature Protocols* **1**, 486–494 (2006).
- Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. H. Werren for providing laboratory strains of *Nasonia* and *Trichomalopsis* and D. Wheeler for discussion on the gene expression data. G. Amdam and J. Liebig allowed us to use their research facilities for conducting the gene expression studies, knockdown experiments and gas chromatography analyses. We thank D. D. McKenna and R. S. Peters for comments on an earlier draft of this paper. O.N. acknowledges the Alexander von Humboldt foundation for a Feodor Lynen postdoctoral research stipend. J.B. and T.S. were supported by the Excellence Initiative of the German Research Foundation (GSC-4, Spemann Graduate School). Parts of this research were supported by the German Research Foundation (DFG) grant RU 717/10-1 to J.R.

Author Contributions Authors J.B., J.D.G., J.R. and O.N. contributed equally to this work. T.S. initiated the study. J.B., J.G., J.D.G., J.R., O.N. and T.S. conceived the experiments. D.P. conducted the bioassays. C.H., J.B., J.R. and T.S. conducted the chemical analyses. A.K.J., C.H., J.B., J.D.G., J.G., N.S.M. and O.N. performed the QTL analyses and knockdown experiments. J.D.G. and O.N. performed the transcriptional analyses. J.G., J.R., O.N. and T.S. provided material and resources. O.N. was responsible for the comparative sequence analysis and took the lead in writing the manuscript. J.R., O.N. and T.S. were the main contributors to the writing of the manuscript.

Author Information The sequences reported in this article are deposited in GenBank under accession numbers FN429934–FN429952, FN430419 and HE962018–HE962021. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.N. (o.niehuus.zfmk@uni-bonn.de).

Ecosystem resilience despite large-scale altered hydroclimatic conditions

Guillermo E. Ponce Campos^{1,2}, M. Susan Moran¹, Alfredo Huete³, Yongguang Zhang¹, Cynthia Bresloff², Travis E. Huxman⁴, Derek Eamus³, David D. Bosch⁵, Anthony R. Buda⁶, Stacey A. Gunter⁷, Tamara Heartsill Scalley⁸, Stanley G. Kitchen⁹, Mitchel P. McClaran¹⁰, W. Henry McNab¹¹, Diane S. Montoya¹², Jack A. Morgan¹³, Debra P. C. Peters¹⁴, E. John Sadler¹⁵, Mark S. Seyfried¹⁶ & Patrick J. Starks¹⁷

Climate change is predicted to increase both drought frequency and duration, and when coupled with substantial warming, will establish a new hydroclimatological model for many regions¹. Large-scale, warm droughts have recently occurred in North America, Africa, Europe, Amazonia and Australia, resulting in major effects on terrestrial ecosystems, carbon balance and food security^{2,3}. Here we compare the functional response of above-ground net primary production to contrasting hydroclimatic periods in the late twentieth century (1975–1998), and drier, warmer conditions in the early twenty-first century (2000–2009) in the Northern and Southern Hemispheres. We find a common ecosystem water-use efficiency (WUE_e : above-ground net primary production/evapotranspiration) across biomes ranging from grassland to forest that indicates an intrinsic system sensitivity to water availability across rainfall regimes, regardless of hydroclimatic conditions. We found higher WUE_e in drier years that increased significantly with drought to a maximum WUE_e across all biomes; and a minimum native state in wetter years that was common across hydroclimatic periods. This indicates biome-scale resilience to the interannual variability associated with the early twenty-first century drought—that is, the capacity to tolerate low, annual precipitation and to respond to subsequent periods of favourable water balance. These findings provide a conceptual model of ecosystem properties at the decadal scale applicable to the widespread altered hydroclimatic conditions that are predicted for later this century. Understanding the hydroclimatic threshold that will break down ecosystem resilience and alter maximum WUE_e may allow us to predict land-surface consequences as large regions become more arid, starting with water-limited, low-productivity grasslands.

Increased aridity and persistent droughts are projected in the twenty-first century for most of Africa, southern Europe and the Middle East, most of the Americas, Australia and South East Asia¹. This is predicted to change vegetation productivity markedly across ecosystems from grasslands to forests^{2,4,5} and affect societal needs for food security and basic livelihood⁶. However, model predictions of productivity responses only provide the most-likely scenarios of the impact of climate change, and few experiments have focused on how anticipated changes in precipitation might be generalized across terrestrial ecosystems. Long-term measurements of natural variability in field settings, supported by manipulative experiments, are considered the best approach for determining the effect of prolonged drought on vegetation productivity^{6,7}.

In field experiments, vegetation productivity is generally measured as the above-ground net primary production (ANPP, or total new organic matter produced above-ground during a specific interval⁸), and vegetation response to changes in precipitation is quantified as rain-use efficiency (RUE), defined as the ratio of ANPP to precipitation over a defined season or year⁹. Using this approach, continental-scale patterns of RUE have been reported for extended periods in the late twentieth century¹⁰. Ecosystem water-use efficiency (WUE_e : ANPP/evapotranspiration¹¹) provides further insight into the ecological functioning of the land surface, in which evapotranspiration is calculated as precipitation minus the water lost to surface runoff, recharge to groundwater and changes to soil water storage¹² (see Methods). Here we compare the functional responses of RUE and WUE_e to local changes in precipitation to document ecosystem resilience—the capacity to absorb disturbances and retain the same function, feedbacks and sensitivity¹³—during altered hydroclimatic conditions.

The objective was to determine how ANPP across biomes responded to altered hydroclimatic conditions forced by the contemporary drought in Southern and Northern Hemispheres from 2000–2009. Measurements made at 12 US Department of Agriculture (USDA) long-term experimental sites in the conterminous United States and Puerto Rico, and 17 similar sites in the Australian continent over a range of precipitation regimes (termed USDA_{00–09} and Australia_{01–09}, respectively). To contrast productivity under altered hydroclimatic conditions with precipitation variability in the late twentieth century, we compared results from the 2000–2009 period with similar analysis of measurements made during the period from 1975–1998 (ref. 10). The latter measurements were primarily from Long-term Ecological Research (LTER) locations, with 14 sites—12 in North America, 2 in Central and South America—hereafter referred to as the LTER_{75–98} data set. For a subset of the LTER_{75–98} sites, ANPP measurements were continued during the period from 2000–2009 (termed LTER_{00–09}), and these were used for further validation of the results (see Supplementary Information and Supplementary Table 1).

The warm drought during the early twenty-first century in the United States, Europe and Australia has been recognized as a considerable change from the climatological variability of the late twentieth century^{1,14}. Globally, 2000–2009 ranked as the ten warmest years of the 130-year (1880–2009) record¹⁵. Global annual evapotranspiration increased on average by 7.1 mm yr^{−1} decade^{−1} from 1982–1997, and after that, remained at a plateau through 2008, thus revealing the

¹USDA ARS Southwest Watershed Research, Tucson, Arizona 85719, USA. ²Soil, Water & Environmental Sciences, University of Arizona, Tucson, Arizona 85721, USA. ³Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, New South Wales 2007, Australia. ⁴Ecology & Evolutionary Biology, University of California, Irvine, California, USA and Center for Environmental Biology, University of California, Irvine, California 92697, USA. ⁵USDA ARS Southeast Watershed Research Laboratory, Tifton, Georgia 31793, USA. ⁶USDA ARS Pasture Systems & Watershed Management Research Unit, University Park, Pennsylvania 16802, USA. ⁷USDA ARS Southern Plains Range Research Station, Woodward, Oklahoma 73801, USA. ⁸USDA FS International Institute of Tropical Forestry, Rio Piedras 00926, Puerto Rico. ⁹USDA FS Rocky Mountain Research Station Shrub Sciences Laboratory, Provo, Utah 84606, USA. ¹⁰School of Natural Resources & the Environment, University of Arizona, Tucson, Arizona 85721, USA. ¹¹USDA FS Southern Research Station, Asheville, North Carolina 28806, USA. ¹²USDA FS Pacific Southwest Research Station, Arcata, California 95521, USA. ¹³USDA ARS Rangeland Resources Research Unit, Fort Collins, Colorado 80526, USA. ¹⁴USDA ARS Jornada Experimental Range & Jornada Basin Long Term Ecological Research Program, New Mexico State University, Las Cruces, New Mexico 88012, USA. ¹⁵USDA ARS Cropping Systems & Water Quality Research Unit, Columbia, Missouri 65211, USA. ¹⁶USDA ARS Northwest Watershed Research Center, Boise, Idaho 83712, USA. ¹⁷USDA ARS Grazinglands Research Laboratory, El Reno, Oklahoma 73036, USA.

impact of the drought on this important Earth surface process¹⁶. In the United States, heatwaves in 2005, 2006 and 2007 broke all-time records for high maximum and minimum temperatures, and drier than average conditions were reported for more than 50% of the conterminous United States in 2000–2002 and 2006–2007 (ref. 17). In Australia, the widespread six-year drought from 2001 to 2007 is considered the most severe in the nation's history¹⁸. The mean Palmer Drought Severity Index (PDSI; see Methods) for USDA and Australian sites decreased significantly ($P < 0.002$) from 1980–1999 to 2000–2009 (USDA) and to 2001–2009 (Australia), declining from -0.06 to -0.81 and from 0.09 to -1.34 , respectively, where a reduction in the PDSI indicates an increase in aridity. Furthermore, warm-season temperatures at USDA and Australian sites during the 2000–2009 periods were significantly higher ($P < 0.014$) than 1980–1999 averages, warming by 0.32 and 0.44 °C, respectively.

The enhanced vegetation index (EVI)¹⁹ satellite observations from the Moderate Resolution Imaging Spectroradiometer (MODIS) were integrated annually (termed iEVI) as an empirical proxy for ANPP at USDA_{00–09} and Australia_{01–09} sites (see Methods). Several publications suggest that this is a robust approximation of collective plant behaviour²⁰, and in this study, we quantified the accuracy of this relationship for the biomes and precipitation patterns. *In situ* estimates of ANPP made with conventional field methods (ANPP_G) between 2000 and 2009 were compiled for ten sites across the United States (see Supplementary Information and Supplementary Table 2) and compared with iEVI measurements for the same site and year (Fig. 1). A log–log regression equation was used to estimate ANPP from iEVI values (ANPP_S), in which $\text{ANPP}_S = 51.42 \times \text{iEVI}^{1.15}$ (Fig. 1).

The response of plant production to precipitation during the contemporary hydroclimatic conditions of prolonged warm drought showed strong agreement with the ANPP/precipitation relations reported during the late twentieth century¹⁰ (Fig. 2a). The lowest mean RUE (that is, the slope of the ANPP/precipitation relationship) reported for biomes with the highest mean precipitation can be explained largely (although not completely¹⁰) by the rain water that is not available for plant production owing to runoff, groundwater recharge and increased soil water storage. Thus, the increase in water available for vegetation production with increasing precipitation is partially consumed by non-biological components of the hydrological cycle (that is, runoff and deep drainage). This becomes apparent when production was plotted as a function of evapotranspiration: the mean

ecosystem water-use efficiency (WUE_m) was constant across the entire precipitation gradient (Fig. 2b). Furthermore, there were no significant differences among WUE_m for the three data sets ($P > 0.05$ per homogeneity of regression slope test). Combined, this indicated that all biomes retained their intrinsic sensitivity to water availability during prolonged, warm drought conditions. This suggests that the rules governing how species are organized in terms of their tolerance of hydrological stress are robust despite extended perturbation by low precipitation.

When water limitations at each site were most severe (for the driest years in each multi-year record), a maximum ecosystem WUE_m (WUE_x) across all biomes was revealed for each of the three data sets (Fig. 3a). The WUE_x was significantly higher for the Australia_{01–09} sites (PDSI = -1.34) than for the LTER_{75–98} and USDA_{00–09} sites (PDSI = -0.81 , respectively) ($P < 0.05$, Fig. 3a, inset). This indicates a cross-biome sensitivity to prolonged warm drought

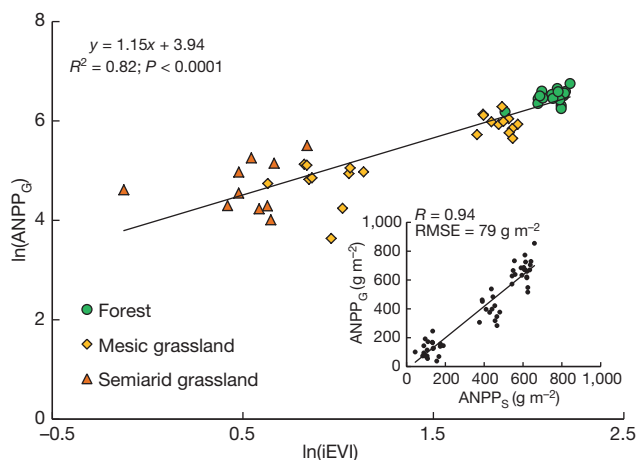


Figure 1 | Relationship between ANPP and iEVI. Relationship between *in situ* estimates of above-ground net primary production (ANPP_G) and iEVI derived from MODIS data (2000–2009) for ten sites across several biomes (see Supplementary Information and Supplementary Table 2). The solid line represents the linear regression used to estimate ANPP from iEVI (ANPP_S), in which $\text{ANPP}_S = 51.42 \times \text{iEVI}^{1.15}$. The inset shows the correlation between estimates of ANPP_S and ANPP_G for the ten sites over several years.

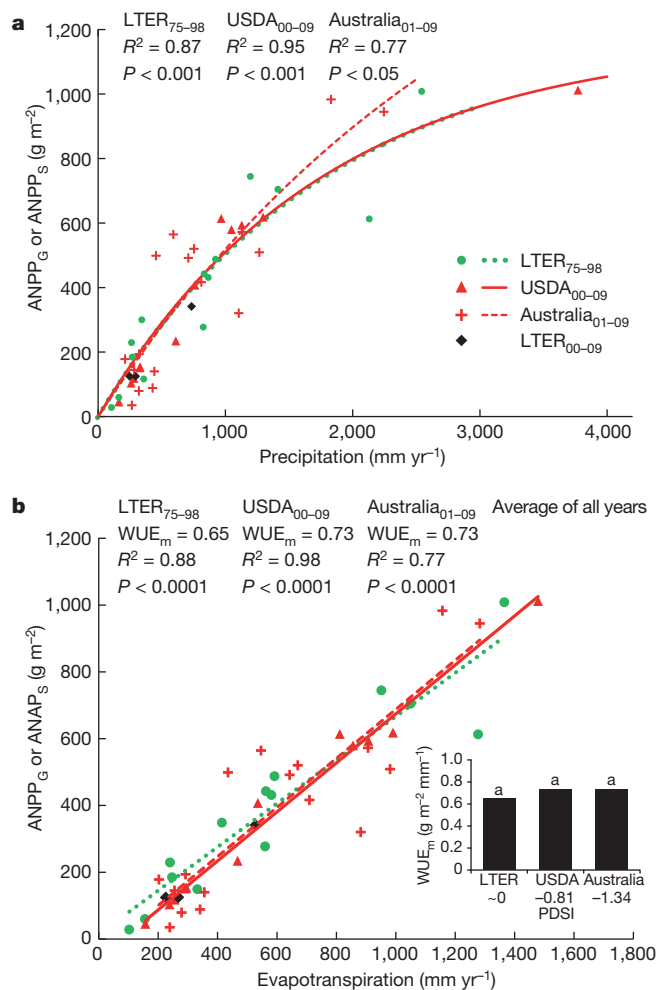


Figure 2 | Cross-biome sensitivity to precipitation during altered hydroclimatic conditions. a, b, Relationship of plant production to precipitation (a) and evapotranspiration (b) across precipitation regimes during the late twentieth century (LTER_{75–98}, green) and during altered hydroclimatic conditions characterized by prolonged, warm drought (USDA_{00–09} and Australia_{01–09}, red), showing significant coefficients of determination in regressions for each data set. Symbols represent the mean values for each site over multi-year study periods. Three LTER sites with *in situ* estimates of ANPP_G during 2000–2009 (black) were included for qualitative validation of results with ANPP_S. The inset in b illustrates differences in mean water-use efficiencies (WUE_m ; the slope of the ANPP/evapotranspiration relationship) across hydroclimatic conditions, in which PDSI ranged from ~ 0 to -1.34 . Columns labelled with the same letter are not significantly different ($P > 0.05$).

where ecosystems sustain productivity in the driest years by increasing their WUE_e . The increase in cross-biome WUE_x with declining PDSI suggests that most biomes were primarily water limited during the driest years of the early twenty-first century drought, and this overshadowed limitations imposed by other resources even at high-productivity sites.

As a test of ecosystem resilience, a similar comparison was made for the wettest years during the mid- to late-drought (2003–2009) and compared to the results for the wettest years during the earlier hydroclimatic conditions from 1975–1998. For the wettest years in both periods, we found a minimum value (WUE_n) that was common to all biomes and similar across both hydroclimatic periods (Fig. 3b). The finding that WUE_n did not vary ($P > 0.05$) across different hydroclimatic periods indicates a cross-biome capacity to respond to high annual precipitation, even during periods of warm drought. The decrease from maximum to minimum WUE_e ranged from 14% (for the USDA_{00–09} and LTER_{75–98} data sets) to 35% (for the Australia_{01–09} data set) and is thought to occur through additional resource constraints that come into play in wet years, including light and nutrient limitations¹⁰. However, it may also be true that the mechanistic relationship between the two time periods was not consistent, where shifts

in species composition as a result of contemporary drought influenced this landscape-scale process.

The ability of plants to increase WUE_x and retain historic WUE_n during altered hydroclimatic conditions suggests that the factors controlling these two processes are different with respect to how the climate and vegetation assemblage are changing. During the driest years, there was a cross-biome adjustment in WUE_e that increased with drought intensity, thus sustaining production at near late-twentieth-century levels during prolonged drought. In the wettest years, the sites exhibited an ability to absorb the disturbances associated with the early twenty-first century drought and retained the same sensitivity of ANPP to water availability across both hydroclimatic periods. These different responses to precipitation extremes may be due to changes in vegetation structure and function, and to plant–soil feedbacks that are not captured in the integrated analysis of either RUE or WUE_e . These must be considered in a full assessment of ecosystem vulnerability or resistance to change.

In this study, ecosystem resilience was measured as the capacity of ecosystems to absorb disturbances associated with the early twenty-first century drought and retain late-twentieth-century sensitivity of ANPP to high annual water availability. Our analyses suggest an intrinsic sensitivity of plant communities to water availability, and a shared capacity to tolerate low annual precipitation but also to respond to high annual precipitation. These findings provide a conceptual model of ecosystem resilience at the decadal scale during the altered hydroclimatic conditions that are predicted for later this century¹ (Fig. 4). During the driest years, the high-productivity sites became water limited to a greater extent resulting in higher WUE_e similar to that encountered in less productive, more arid ecosystems. It follows that when all ecosystems are primarily water limited, a cross-biome maximum WUE_e will be reached (WUE_x) that cannot be sustained with further reductions in water availability. Furthermore, we predict that as cross-biome WUE_e reaches that maximum WUE_x value, WUE_n will approach WUE_x because production will be limited largely by water supply and less so by nutrients and light (Fig. 4).

With continuing warm drought, the single linear ANPP/evapotranspiration relation that forms the common cross-biome WUE_e would collapse as biomes endure the significant drought-induced mortality

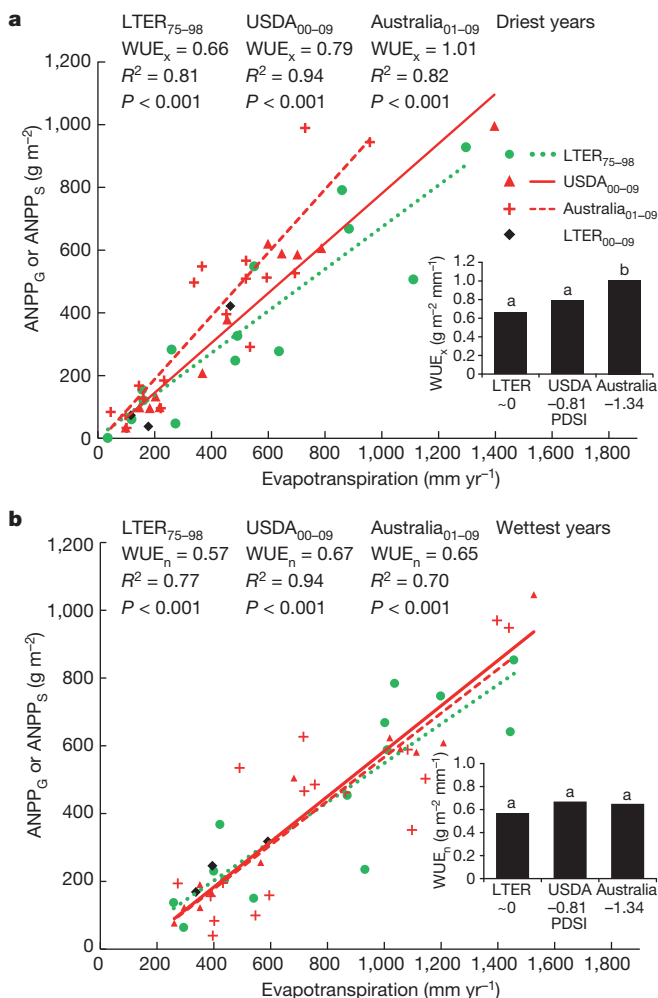


Figure 3 | Ecosystem resilience across biomes and hydroclimatic conditions. **a, b**, Maximum (WUE_x) (**a**) and minimum (WUE_n) (**b**) water-use efficiency (slope of the ANPP/evapotranspiration) in the driest and wettest years, respectively, based on all sites for each data set, plus the three LTER_{00–09} validation sites. The insets illustrate the differences in WUE_x (**a**) and WUE_n (**b**) with mean PDSI for the study periods and locations. Columns labelled with the same letter are not significantly different ($P > 0.05$) across hydroclimatic conditions.

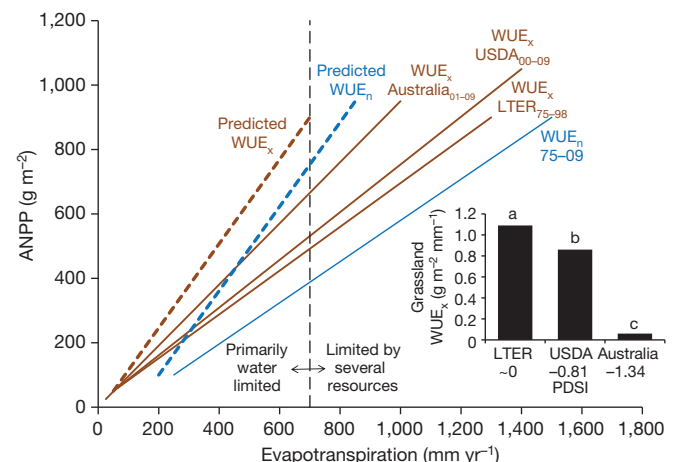


Figure 4 | A conceptual model of ecosystem resilience during altered hydroclimatic condition. A summary of WUE_e results in this study (solid lines), overlain with the predicted behaviour of WUE_x (brown dashed line) and WUE_n (blue dashed line) along a continuum of sites limited primarily by water and by other resources with an arbitrary distinction made here at evapotranspiration = 700 mm yr⁻¹ for illustration only (black dashed line). Predictions are based on forecasts of continuing warm drought¹. The inset illustrates the decrease in WUE_x with PDSI for subsets of the LTER_{75–98} ($n = 4$), USDA_{00–09} ($n = 5$) and Australia_{01–09} ($n = 2$) data sets limited to grassland sites, where columns labelled with the same letter are not significantly different ($P > 0.05$).

that has been extensively documented over the past decade^{2,5}. This loss of resilience associated with dieback would probably occur first for ecosystems that respond most rapidly to precipitation variability (that is, grasslands²¹). Thus, the cross-biome ANPP/evapotranspiration relation would become nonlinear as WUE_x and WUE_n approached zero for the most water-limited, low-productivity sites, whereas WUE_e values would be less affected in the high-productivity sites. Subsets of the LTER_{75–98} ($n = 4$), USDA_{00–09} ($n = 5$) and Australia_{01–09} ($n = 2$) data sets limited to grassland sites were used to corroborate this prediction (Fig. 4, inset). During this study period, grassland WUE_x decreased with increasing aridity (decreasing PDSI), indicating a decreasing resilience with prolonged warm drought in these biomes, as predicted. This suggests that these systems are closer to a threshold which, when crossed, will result in biome reorganization.

Here we quantified the effect of the early twenty-first century drought on ecosystem productivity and resilience across many sites on two continents. Cross-biome capacities and sensitivities of production were maintained through prolonged warm drought by increases of WUE_e during the driest years and resilience during wet years indicated by a common WUE_e across both hydroclimatic periods. The conclusions are particularly compelling because they are based on measurements across several biomes with comparisons of multi-year periods of altered hydroclimatic conditions. These findings were extended to predictions that if warm drought continues, considerable mortality (particularly in low-productivity grasslands that are most sensitive to water availability) may threaten ecosystem resilience across biomes given the substantial changes in ecosystem structure. The emergence of these patterns at the spatial and temporal scale at which they were derived requires investigation of the supporting eco-hydrological mechanisms that underlie the complex plant–soil couplings. Spatially, this work represents broad cross-biome behaviour but does not fully represent the complex site-level response to prolonged warm drought. The site-level mechanisms associated with disease, pests, fire, response lags, species replacement and meristem density in forests² and grasslands^{4,21} complicate specific processes maintaining or affecting cross-biome resilience of ecosystem function. Furthermore, there are predictions of a general biogeochemical resetting as increases in carbon dioxide supply affect many plant and soil processes²². Temporally, these predictions of ecosystem resilience were based on behaviour at the scale of a decade or longer, including a period of prolonged warm drought. With careful application of this satellite-based metric, it is possible to continue monitoring cross-biome ecosystem resilience at selected cross-continental sites year-by-year into the future as we develop a greater understanding of the physical and biological mechanisms controlling these patterns.

METHODS SUMMARY

Daily precipitation and temperature were measured at *in situ* stations and represented a homogeneous vegetated area of $\sim 3 \times 3$ km with no major disturbances (for example, fires) during the 2000–2009 period. The total and the mean annual precipitation were computed from daily values over the study period during the hydrological year (October–September for the United States, and May–April for Australia). PDSI values were computed using precipitation, temperature and soil water holding capacity data (see Methods). EVI values were extracted from MODIS images at 250-m spatial resolution for a window size of 9×9 pixels and averaged to one value every 16 days for a time series that was smoothed to obtain iEVI (see Methods). Estimates of annual evapotranspiration were obtained by incorporating annual precipitation and percentages of forested and herbaceous

cover in a model derived from more than 250 catchment-scale measurements from around the world¹².

Full Methods and any associated references are available in the online version of the paper.

Received 6 March; accepted 3 December 2012.

Published online 20 January 2013.

1. Dai, A. Drought under global warming: a review. *WIREs Clim. Change* **2**, 45–65 (2011).
2. Breshears, D. D. *et al.* Regional vegetation die-off in response to global-change-type drought. *Proc. Natl Acad. Sci. USA* **102**, 15144–15148 (2005).
3. Saleska, S. R., Didan, K., Huete, A. R. & da Rocha, H. R. Amazon forests green-up during 2005 drought. *Science* **318**, 612 (2007).
4. Scott, R. L., Hamerlynck, E. P., Jenerette, G. D., Moran, M. S. & Barron-Gafford, G. A. Carbon dioxide exchange in a semidesert grassland through drought-induced vegetation change. *J. Geophys. Res.* **115**, G03026 (2010).
5. Allen, C. D. *et al.* A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259**, 660–684 (2010).
6. Milly, P. C. D. *et al.* Stationarity is dead: whither water management? *Science* **319**, 573–574 (2008).
7. Weltzin, J. F. *et al.* Assessing the response of terrestrial ecosystems to potential changes in precipitation. *Bioscience* **53**, 941–952 (2003).
8. Roxburgh, S. H., Berry, S. L., Buckley, T. N., Barnes, B. & Roderick, M. L. What is NPP? Inconsistent accounting of respiratory fluxes in the definition of net primary production. *Funct. Ecol.* **19**, 378–382 (2005).
9. Le Houérou, H. N. Rain use efficiency: a unifying concept in arid-land ecology. *J. Arid Environ.* **7**, 213 (1984).
10. Huxman, T. E. *et al.* Convergence across biomes to a common rain-use efficiency. *Nature* **429**, 651–654 (2004).
11. Monson, R. K. *et al.* Tree species effects on ecosystem water-use efficiency in a high-elevation, subalpine forest. *Oecologia* **162**, 491–504 (2010).
12. Zhang, L., Dawes, W. R. & Walker, G. R. Response of mean annual evapotranspiration to vegetation changes at catchment scale. *Water Resour. Res.* **37**, 701–708 (2001).
13. Walker, B., Holling, C. S., Carpenter, S. R. & Kinzig, A. Resilience, adaptability and transformability in social–ecological systems. *Ecol. Soc.* **9**, 5 (2004).
14. MacDonald, G. M. Water, climate change, and sustainability in the southwest. *Proc. Natl Acad. Sci. USA* **107**, 21256–21262 (2010).
15. National Oceanic and Atmospheric Administration. US climate division data plots (<http://www.esrl.noaa.gov/psd/data/usclimdivs/>) (2012).
16. Jung, M. *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954 (2010).
17. National Drought Mitigation Center. U.S. drought monitor (<http://drought.unl.edu/MonitoringTools/USDroughtMonitor.aspx>) (2012).
18. Australian Government Bureau of Meteorology. Australia's climate change datasets (<http://www.bom.gov.au/climate/change/datasets/datasets.shtml>) (2011).
19. Huete, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **83**, 195–213 (2002).
20. Running, S. W. *et al.* A continuous satellite-derived measure of global terrestrial primary production. *Bioscience* **54**, 547 (2004).
21. Knapp, A. K. & Smith, M. D. Variation among biomes in temporal dynamics of aboveground primary production. *Science* **291**, 481–484 (2001).
22. Morgan, J. A. *et al.* C4 grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland. *Nature* **476**, 202–205 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements The work was supported in part by the NASA SMAP Science Definition Team under agreement 08-SMAPSDT08-0042, the Australian Research Council (ARC) Discover Project (DP1115479) and the Terrestrial Ecosystem Research Network (TERN) EIF: AusCover. We thank the Australian Bureau of Meteorology for providing the precipitation data. We also thank J. Overpeck, T. McVicar, R. Donohue and M. Walbridge for their input.

Author Contributions G.E.P.C., M.S.M. and A.H. conceived the study, assembled the data and produced the preliminary results. The remaining authors collected and analysed data, and contributed to the interpretation of results. All authors contributed to writing the paper. Statistical analyses were performed by G.E.P.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.E.P.C. (geponce@gmail.com) or M.S.M. (susan.moran@ars.usda.gov).

METHODS

MODIS iEVI. At each USDA_{00–09} and Australia_{01–09} site (see Supplementary Information and Supplementary Table 1), MODIS EVI (MOD13Q1) data with 16-day and 250-m temporal and spatial resolutions, respectively, were acquired from the corresponding MODIS tiles. Sites were selected where vegetation was uniform across space and *in situ* climate records representative of the location could be obtained. We extracted areas of 2.25×2.25 km (9×9 250-m pixels) for each location, yielding 23 16-day image files per year for a total of 230 files for each site. Pixel-based quality assurance control was applied to the time series data to remove low-quality, cloud- and aerosol-contaminated pixels and observations made at large sensor zenith angles ($>30^\circ$). The retained high-quality pixels were averaged to represent the EVI value for that site and 16-day period, resulting in a 10-year EVI time series for each site.

The next step was to use the software tool TIMESAT²³ to smooth the quality-assurance-filtered time series data and standardize the MODIS EVI time series analysis for consistent cross-site comparisons. The TIMESAT filtering option known as the adaptive Savitzky–Golay filter²⁴ was applied over the time series for smoothing the data and suppressing noise by replacing each data value y_i , $i = 1, \dots, n$ by a linear combination of nearby values in a window, where

$$\sum_{j=-n}^n c_j y_{i+j} \quad (1)$$

and the weights were $c_j = 1/(2n + 1)$, and the data value y_i was replaced by the average of the values in the window. The moving-average method preserved the area and mean position of a seasonal peak, but altered both the width and height. The latter properties were preserved by approximating the underlying data value with the value obtained from a least-square fit to a polynomial, rather than the average in the window. For each data value y_i , $i = 1, 2, \dots, n$ a quadratic polynomial was fitted as $f(t) = c_1 + c_2 t + c_3 t^2$ to all $2n + 1$ points in the moving window and the value y_i was replaced with the value of the polynomial at position t_i (ref. 24). The advantage of this method was that it preserved features of the distribution such as relative maxima, minima and width, which were usually ‘flattened’ by other adjacent-based averaging techniques.

To simplify the process of integrating EVI values from TIMESAT and avoid parameterization, we integrated over the entire year for every site. Therefore, the process of integrating EVI to obtain iEVI was based on using the default parameters found when TIMESAT was initiated. After smoothing the series, we proceeded to extract an offset of 0.05 of each 16-day EVI value to reduce effects of soil exposure. Our process was standardized by applying the same procedures to each data set used.

Meteorological data. Daily precipitation and temperature were measured at *in situ* stations associated with the experimental sites. Total annual precipitation (sum of daily precipitation, mm yr^{-1}), mean annual precipitation (MAP) (mean of annual precipitation over the study period, mm yr^{-1}) and mean maximum temperature (mean of average monthly maximum temperature, $^\circ\text{C}$) were computed for the hydrological year, defined as the 12-month period from October–September in the Northern Hemisphere and May–April in the Southern Hemisphere. The warm season was defined as April–September for USDA sites and November–April for Australia sites.

PDSI. The PDSI²⁵ was computed with the Thornthwaite equation²⁶ using a self-calibrating PDSI implementation that automatically calibrated the behaviour of

the index at any location²⁷. The Thornthwaite equation computes potential evapotranspiration (PET) = $16d(10T/I)^a$, in which T is the mean temperature for the month, d is a correction factor that depends on latitude and month, I is an annual thermal index, and a is an empirical factor²⁶. The three main PDSI inputs were monthly rainfall, monthly temperature and soil water holding capacity (SWHC). Rainfall and temperature were obtained from the stations used in this study. We obtained the SWHC data from the NRCS web soil survey (<http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx>) for the USDA sites and the Australian Soils Resource Information System (ASRIS; http://www.asris.csiro.au/index_other.html) for the Australia sites.

The PDSI is a measure of drought and wet spells, in which PDSI = 0 is normal, −3 is moderate drought, −4 is extreme drought, and excess precipitation is represented by a positive PDSI. We obtained the PDSI for the time period from 1980 to 2009 to identify the average drought conditions across the USDA, LTER and Australia sites. On the basis of this site-specific PDSI (see Supplementary Information and Supplementary Table 3) and reports of continental-scale drought extent and severity^{15–18} (summarized in the main text), the period of altered hydroclimatic conditions was determined to be 2000–2009 for USDA sites and 2001–2009 for Australia sites, reflected by the naming convention USDA_{00–09} and Australia_{01–09}, respectively.

Evapotranspiration model. Estimates of evapotranspiration at different biomes were obtained using a model of mean annual evapotranspiration formulated with data from more than 250 catchment-scale measurements from around the world¹². The two-parameter model offers an approach for estimation of mean annual evapotranspiration (ET) on the basis of changes in annual precipitation (P) (mm yr^{-1}) and the percentage of forest cover (f), where

$$ET = \left(f \frac{1 + 2 \frac{1,410}{P}}{1 + 2 \frac{1,410}{P} + \frac{P}{1,410}} + (1-f) \frac{1 + 0.5 \frac{1,100}{P}}{1 + \frac{1,100}{P} + \frac{P}{1,100}} \right) P \quad (2)$$

The model has two portions as depicted in equation (2), with the left side accounting for the fractional forest cover and the right side accounting for the fractional herbaceous plant cover (non-forested). The main advantage of this model over more traditional models is the derivation from data readily available at the catchment scale. For the USDA_{00–09} data set, the information about the percentage of non-forested areas was obtained from contacts at each location. For the Australia_{01–09} data set, estimations of the percentage of non-forested areas were made using Google Earth.

23. Jönsson, P. & Eklundh, L. TIMESAT—a program for analyzing time-series of satellite sensor data. *Comput. Geosci.* **30**, 833–845 (2004).
24. Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639 (1964).
25. Palmer, W. C. *Meteorological Drought* Weather Bureau Res. Paper no.45 (1965).
26. Thornthwaite, C. W. An approach toward a rational classification of climate. *Geogr. Rev.* **38**, 55 (1948).
27. Wells, N., Goddard, S. & Hayes, M. J. A self-calibrating Palmer Drought Severity Index. *J. Clim.* **17**, 2335–2351 (2004).

Stabilization of cooperative virulence by the expression of an avirulent phenotype

Médéric Diard¹, Victor Garcia², Lisa Maier¹, Mitja N. P. Remus-Emsermann¹, Roland R. Regoes², Martin Ackermann³ & Wolf-Dietrich Hardt¹

Pathogens often infect hosts through collective actions: they secrete growth-promoting compounds or virulence factors, or evoke host reactions that fuel the colonization of the host. Such behaviours are vulnerable to the rise of mutants that benefit from the collective action without contributing to it; how these behaviours can be evolutionarily stable is not well understood¹. We address this question using the intestinal pathogen *Salmonella enterica* serovar Typhimurium (hereafter termed *S. typhimurium*), which manipulates its host to induce inflammation, and thereby outcompetes the commensal microbiota^{2,3}. Notably, the virulence factors needed for host manipulation are expressed in a bistable fashion, leading to a slow-growing subpopulation that expresses virulence genes, and a fast-growing subpopulation that is phenotypically avirulent^{4,5}. Here we show that the expression of the genetically identical but phenotypically avirulent subpopulation is essential for the evolutionary stability of virulence in this pathogen. Using a combination of mathematical modelling, experimental evolution and competition experiments we found that within-host evolution leads to the emergence of mutants that are genetically avirulent and fast-growing. These mutants are defectors that exploit inflammation without contributing to it. In infection experiments initiated with wild-type *S. typhimurium*, defectors increase only slowly in frequency. In a genetically modified *S. typhimurium* strain in which the phenotypically avirulent subpopulation is reduced in size, defectors rise more rapidly, inflammation ceases prematurely, and *S. typhimurium* is quickly cleared from the gut. Our results establish that host manipulation by *S. typhimurium* is a cooperative trait that is vulnerable to the rise of avirulent defectors; the expression of a phenotypically avirulent subpopulation that grows as fast as defectors slows down this process, and thereby promotes the evolutionary stability of virulence. This points to a key role of bistable virulence gene expression in stabilizing cooperative virulence and may lead the way to new approaches for controlling pathogens.

Virulence refers to damage done by a pathogen to its hosts⁶. It is thus important to understand how determinants of virulence are regulated, and how virulence evolves. Evolutionary theory offers a framework for understanding how costs and benefits to the pathogen determine the evolution of virulence. In general, competition between different pathogen genotypes within the same host will select for increased host exploitation and increased virulence^{7,8}. A different outcome has been predicted for virulence based on the cooperative secretion of costly compounds^{9–11}. Avirulent mutants that do not contribute to secretion, but reap the benefit, can rise in frequency, and within-host evolution can thus lead to decreased virulence^{12–15}. Given the ease of gene loss and mutation, it seems surprising that cooperative virulence can be maintained.

To address this question, we used *S. typhimurium* as an example of a pathogen that uses costly virulence factors to manipulate the host^{2–5}. *S. typhimurium* expresses the type III secretion system 1 (*ttss-1*, also

known as T3SS-1 and T1) to induce gut inflammation, which is required to overcome competition from the commensal microbiota³. Importantly, only a fraction of the cells in a clonal population of *S. typhimurium* express *ttss-1* (Supplementary Fig. 6), and these cells pay a cost in terms of slow growth *in vitro*⁵ and *in vivo* (Supplementary Figs 7 and 8); high mortality after invasion of the mucosa may inflict further costs⁴. We refer to this subpopulation as T1⁺. A second subpopulation (T1[−]) is phenotypically avirulent with respect to *ttss-1* expression, and does not pay the costs of virulence^{4,5}. Although the molecular mechanism of bistable *ttss-1* expression is not fully understood, here we focus on the consequences of bistable expression for the evolutionary stability of virulence. We have previously interpreted host manipulation by the T1⁺ subpopulation as a cooperative act⁴. Here we directly test whether the induction of inflammation is a cooperative trait that is prone to the invasion of avirulent mutants that do not express a T1⁺ subpopulation, and how this affects within-host evolution of virulence. We use mathematical modelling to derive explicit predictions, which we then tested using experimental evolution in mouse infection models.

We modelled the outcome of competition in the mouse gut between commensal gut bacteria, wild-type cooperating *S. typhimurium*, and genetically avirulent *S. typhimurium*. The mathematical model was adapted from previous work^{4,16} and incorporated empirically measured costs^{4,5} and benefits² for *S. typhimurium* resulting from *ttss-1*-mediated inflammation (see Supplementary Material and Supplementary Figs 1–5). In our model, wild-type *S. typhimurium* quickly reaches numerical dominance over commensal bacteria, and forms a stable population in the infected host for at least 10 days (Fig. 1a and Supplementary Fig. 2). This is in line with typical infection data. If we instead model an avirulent mutant of *S. typhimurium* that does not express a T1⁺ subpopulation, this mutant cannot increase in frequency and is lost from the host (Fig. 1b). However, if the avirulent mutant emerges through spontaneous mutation in a population of wild-type *S. typhimurium*, it can benefit from the inflammation induced by the T1⁺ subpopulation formed by the wild type, and increases in frequency (Fig. 1c). This behaviour characterizes these mutants as defectors. Therefore, the first prediction of our model is that host manipulation by *S. typhimurium* is a cooperative trait that can be exploited by avirulent defectors that benefit from cooperative virulence without contributing to it.

How cooperative virulence can be maintained in the face of defectors is unclear. The phenotypically avirulent T1[−] subpopulation, which grows as fast as avirulent mutants⁵, is expected to slow down the rise of defectors. Indeed, in our model, wild-type *S. typhimurium* populations comprising a substantial fraction of T1[−] cells (for example, 63%; Fig. 1c and Supplementary Fig. 3) can remain dominant over defectors for more than 300 h, thus giving genetically virulent wild-type *S. typhimurium* ample time for growth and transmission. If we model an alternative scenario in which the T1[−] fraction is

¹Institute of Microbiology, ETH Zurich, Wolfgang-Pauli-Str. 10, 8093 Zurich, Switzerland. ²Institute of Integrative Biology, ETH Zurich, Universitätsstr. 16, 8092 Zurich, Switzerland. ³Department of Environmental Systems Science, ETH Zurich, and Department of Environmental Microbiology Eawag, Ueberlandstr. 133, PO Box 611, 8600 Dübendorf, Switzerland.

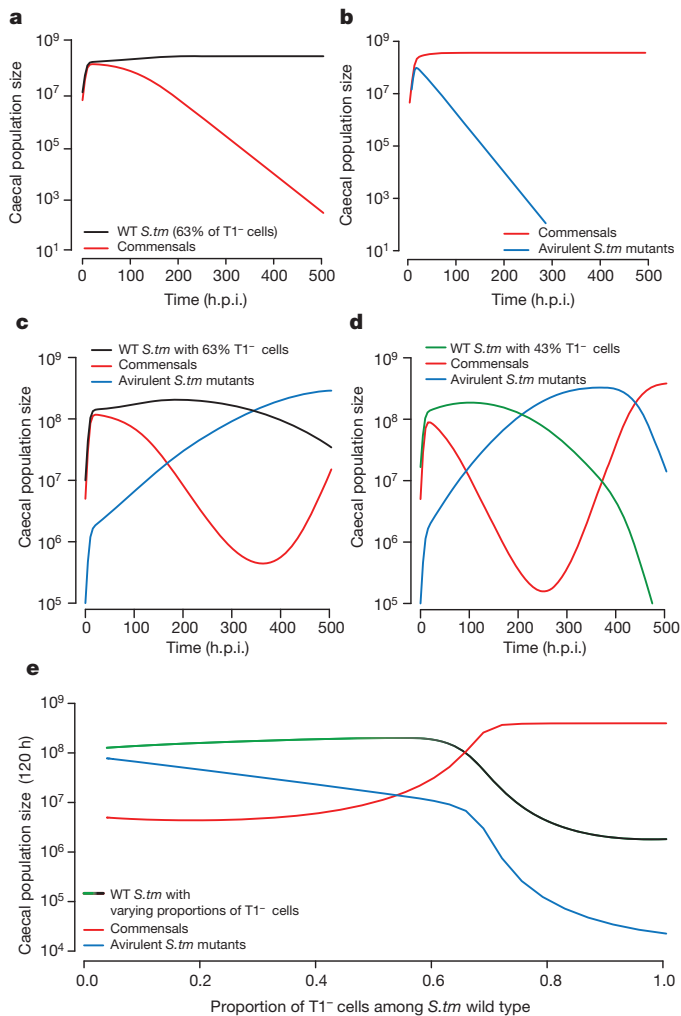


Figure 1 | Mathematical modelling predicts that within-host evolution leads to decreased virulence in *S. typhimurium* infections. The model describes the dynamics of competition between wild-type *S. typhimurium* (*S. tm*), commensal gut microbiota, and genetically avirulent *S. typhimurium* mutants. **a**, Wild-type (WT) *S. typhimurium* outcompete commensals. h.p.i., hours post infection. **b**, Avirulent *S. typhimurium* mutants cannot outcompete commensals. **c**, Avirulent *S. typhimurium* benefit from wild-type *S. typhimurium* during co-infection. **d**, A decreased $T1^-$ subpopulation size accelerates the rise of avirulent *S. typhimurium* mutants. **e**, Fine-tuning of bistable *ttss-1* expression is essential to mediate between the ability to outcompete commensals and stability against defectors: if the proportion of $T1^-$ cells is too high, *S. typhimurium* cannot elicit inflammation and outcompete commensals; if it is too low, avirulent defectors rise quickly. For details, see Supplementary Information and Supplementary Figs 1–5.

smaller (for example, 43%, which is close to values obtained through de-regulation of *ttss-1* (ref. 5); Fig. 1d and Supplementary Fig. 5) we find that the rise of defectors is accelerated. This leads, in our model, to a rapid cessation of inflammation, regrowth of commensals, and the loss of *S. typhimurium* from the gut (Fig. 1d, e and Supplementary Fig. 5). The second prediction of the model is thus that the rise of avirulent defectors can be limited through bistable expression of *ttss-1* and the resulting formation of the phenotypically avirulent $T1^-$ subpopulation.

We then used experimental evolution in mouse infections to test these predictions. We worked with streptomycin-treated mice infected with *S. typhimurium* strain SL1344 *ssaV*^{17,18}; we refer to this strain as wild type, and denote it as *S. typhimurium**. *S. typhimurium** efficiently colonized the intestine within 2 days post infection (d.p.i.) (Fig. 2a) and triggered lasting gut inflammation (Fig. 2). Next, we asked whether the infected mice would contain genetically avirulent mutants, and whether these mutants would behave as defectors, as predicted by our model. At 2 d.p.i., a small proportion of *S. typhimurium* clones isolated from infected mice did not express *ttss-1* based on an assay for the presence of the *ttss-1* effector SipC (Fig. 2d, Supplementary Fig. 9a, b and Supplementary Information); at 10 d.p.i., the frequency of these clones was significantly higher (Fig. 2d). To characterize these mutants, we re-sequenced clones isolated from

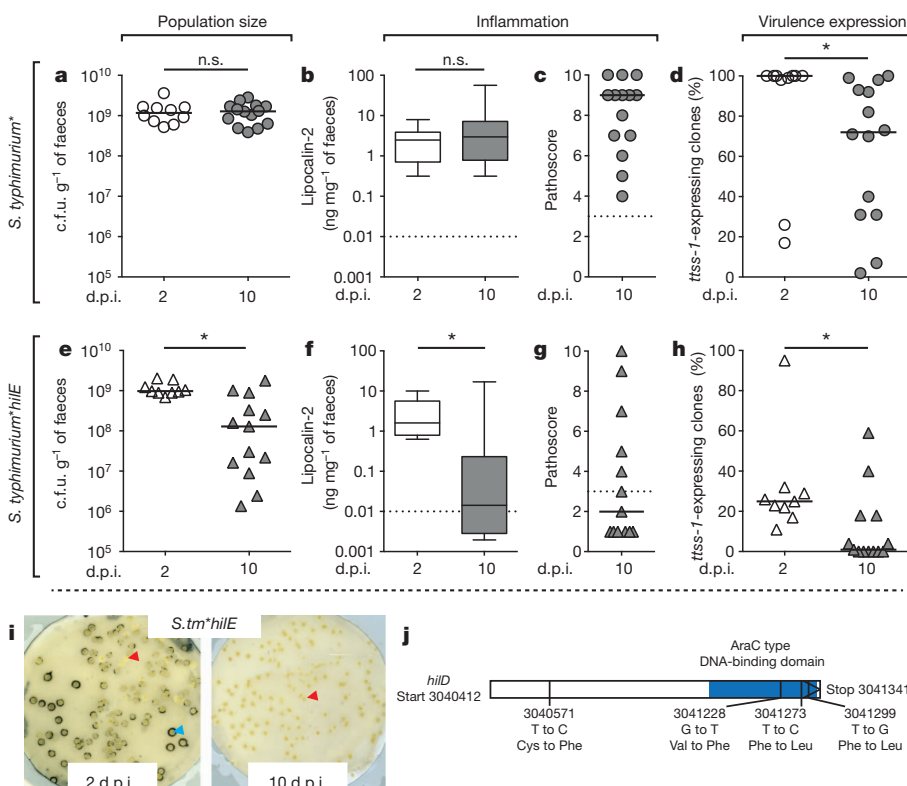


Figure 2 | Infection experiments confirm that avirulent mutants rise during within-host evolution of *S. typhimurium.** **a–i**, Mice were infected with either *S. typhimurium** (60% $T1^-$ subpopulation; **a–d**) or *S. typhimurium***hilE* (30% $T1^-$ subpopulation; **e–i**). We analysed faecal shedding (**a**, **e**), the inflammation marker lipocalin-2 (**b**, **f**; whiskers denote minimum and maximum), mucosal inflammation (**c**, **g**), and the proportion of clones retaining the ability to express *ttss-1* (**d**, **h**, **i**) in faeces at days 2 (white symbols) and 10 (grey symbols) after infection. Dotted lines in **b**, **c**, **f** and **g** indicate typical values of healthy controls. * $P < 0.05$, Mann–Whitney *U* test; n.s., not significant. Blue and red arrowheads in **i** denote SipC⁺ and SipC[−] clones, respectively. **j**, Single nucleotide polymorphisms in *hilD* from re-isolated clones not expressing *ttss-1*.

the faeces of the 10 d.p.i. experiments (Fig. 2, Supplementary Table 1 and Supplementary Figs 10–12). All *ttss-1*-deficient isolates from independent animal experiments contained mutations in *hilD* (Fig. 2d, j and Supplementary Table 1; equivalent observations were made using *S. typhimurium* strains that had no *ssaV* mutation and thus a functional type III secretion system 2; Supplementary Figs 13 and 14). *HilD* is a master-regulator of *ttss-1* expression¹⁹, and the mutations in *hilD* abolished *ttss-1* expression and the ability to trigger inflammation (Supplementary Figs 9c, 15 and 16).

We then tested whether the *hilD* mutants are defectors. We introduced the *hilD* mutation into a wild-type background to isolate it from other mutations that occurred during experimental evolution, resulting in strain *S. typhimurium***hilD*. *S. typhimurium***hilD* was unable to trigger inflammation or sustain a stable intestinal population size for 10 days of infection (Fig. 3a, b and Supplementary Fig. 17). However, it could benefit from *S. typhimurium** during co-infection, and overgrew *S. typhimurium** by a factor of 10 at 3 d.p.i. (Supplementary Fig. 17a); by day 10, shedding had declined by about ten-fold (Fig. 3a), and mucosal inflammation was resolved in most mice (Fig. 3b and Supplementary Fig. 17b). *S. typhimurium***hilD* thus showed the hallmarks of a defector: although unable to trigger inflammation on its own, it could benefit from the inflammation induced by the wild type, leading ultimately to a decline of the *Salmonella* population (Fig. 3a and Supplementary Fig. 17c). This outcome established that the virulence of *S. typhimurium* is a cooperative trait that is prone to invasion by avirulent defectors, and that mutations in *hilD* produce defectors.

We then tested the second prediction of the model: that the relative stability of *S. typhimurium* against defectors crucially depends on the phenotypically avirulent T1[−] subpopulation. To analyse the effects of

varying the size of the T1[−] subpopulation, we constructed a *hilE* mutant of *S. typhimurium** (refs 5, 20). *S. typhimurium***hilE* showed a lower proportion of the T1[−] individuals than *S. typhimurium**, both *in vitro* and *in vivo*⁵ (Supplementary Fig. 6), and we thus expected it to be more vulnerable to the overgrowth of avirulent defectors (Fig. 1d). When introduced into mice, this strain efficiently colonized the intestine within 2 d.p.i. (Fig. 2e) and also induced gut inflammation (Fig. 2f). However, *S. typhimurium***hilE* caused only short-lived infections. By 10 d.p.i., the *Salmonella* population had sharply declined (Fig. 2e) and gut inflammation had subsided (Fig. 2f, g).

We then asked whether the decline of the *S. typhimurium***hilE* population was caused by rapid overgrowth of avirulent defectors during the infection. We found that *ttss-1*-deficient defectors rose quicker and reached higher levels in mice infected with *S. typhimurium***hilE* than in those infected with *S. typhimurium** (Fig. 2h). Re-sequencing of two avirulent clones isolated at 10 d.p.i. from *S. typhimurium***hilE*-infected mice showed that these clones also carried mutations in *hilD* (Supplementary Table 1 and Fig. 2j), and competition experiments with reconstructed genotypes showed that *S. typhimurium***hilE* was more sensitive than wild type to the invasion of *hilD* mutants benefiting from inflammation (Fig. 4 and Supplementary Figs 18–20). These results show that the phenotypically avirulent T1[−] subpopulation is important for delaying the overgrowth by avirulent defectors; if the size of the T1[−] subpopulation is decreased genetically, defectors rise so quickly that the infection is very

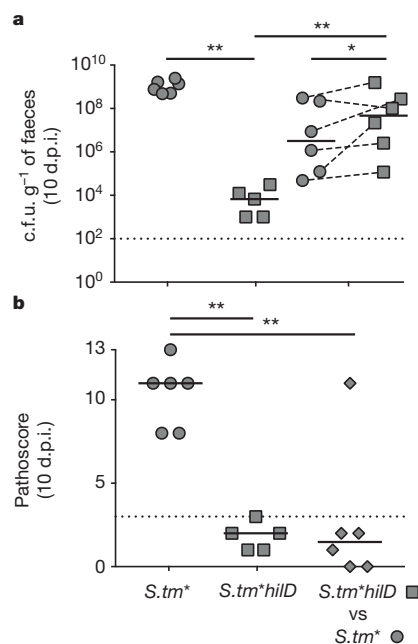


Figure 3 | *S. typhimurium** virulence is a cooperative trait, and avirulent mutants are defectors that exploit host inflammation without contributing to it. Mice were infected with *S. typhimurium**, avirulent *S. typhimurium***hilD*, or 1:1 mixtures of *S. typhimurium** and *S. typhimurium***hilD*. **a**, Population sizes at 10 d.p.i. In co-infections, *S. typhimurium***hilD* benefits from *S. typhimurium**, the total population size reaches intermediate levels, and, within most hosts, *S. typhimurium***hilD* outnumbers wild-type *S. typhimurium**. **b**, The presence of avirulent *S. typhimurium***hilD* leads to a reduced pathoscore, and thus to improved 'health' of the host. * $P < 0.05$, ** $P < 0.01$; one-tailed Wilcoxon matched-pairs signed rank test on paired data (dashed lines in **a**) and Mann–Whitney *U* test (**a**, **b**). Dotted lines indicate the detection limit (**a**) or the threshold between non-inflamed and inflamed caecal tissue (**b**).

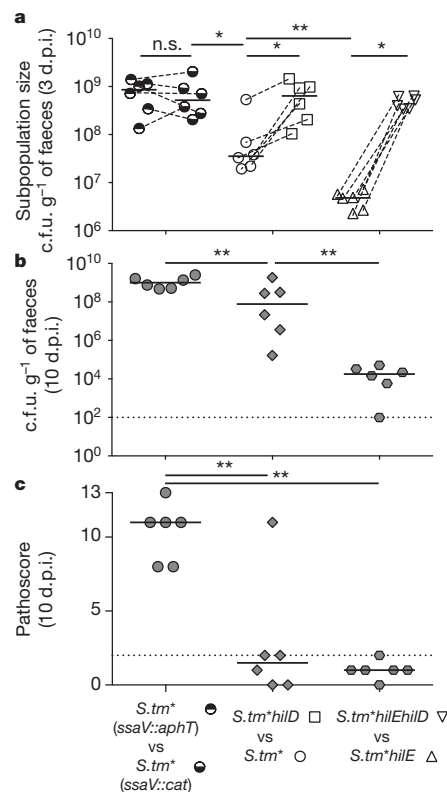


Figure 4 | The phenotypically avirulent T1[−] population slows down the rise of avirulent defectors, and thus stabilizes cooperative virulence. Mice were co-infected with mixtures of either *S. typhimurium** or *S. typhimurium***hilE* and their respective *hilD* derivatives. Co-infections with differently marked *S. typhimurium** strains served as a control. **a**, *hilD* mutants rise more quickly in an *S. typhimurium***hilE* than an *S. typhimurium** background. **b**, The rise of avirulent defectors leads to a decrease in the total *Salmonella* population size. This decrease is stronger in the *S. typhimurium***hilE* background. **c**, The rise of avirulent defectors leads to reduced pathology. * $P < 0.05$, ** $P < 0.01$; one-tailed Wilcoxon matched-pairs signed rank test on paired data (dashed lines in **a**) and Mann–Whitney *U* test (**a**–**c**). Dotted lines indicated the detection limit (**b**) or the threshold between non-inflamed and inflamed caecal tissue (**c**).

short-lived, and the total faecal output of the virulent genotype is markedly reduced.

These findings provide an explanation for the adaptive value of bistability in *S. typhimurium* virulence gene expression and its importance for the maintenance of cooperative virulence. We had previously discussed bistability in virulence gene expression using a more general model of self-destructive cooperation⁴, in which the phenotypically virulent subpopulation is killed completely. Our specific model and the experimental results presented here indicate that in *S. typhimurium*, the virulent T1⁺ subpopulation is not completely killed, but that the growth deficit⁵ and partial killing in the mucosa⁴ render it vulnerable to the overgrowth by avirulent defectors. The phenotypically avirulent T1⁻ subpopulation formed through bistable virulence gene expression delays this process by growing as fast as defectors⁵. This delay could be crucial for the evolutionary stability of cooperative virulence in *S. typhimurium*: although wild-type *S. typhimurium* will eventually be replaced by avirulent mutants within each host, they could prevail in the long run through between-host evolution^{21,22}. Similar dynamics can be expected in other pathogens in which virulence gene expression is costly and exhibits phenotypic heterogeneity in clonal populations. Our results also reveal an untapped potential for combating pathogens: they predict that avirulent defectors protect the host from prolonged infection by virulent strains, and might help to minimize the risk of transmission. The deliberate administration of strains that behave as defectors could thus offer a new strategy for controlling infections. The development of such alternative treatment strategies is particularly important in the face of the current antibiotics crisis²³.

METHODS SUMMARY

Infections were performed in streptomycin-treated C57BL/6 mice using 5×10^7 bacteria. *S. typhimurium* loads and lipocalin-2 were respectively quantified by plating and ELISA of faeces. At 10 d.p.i., mice were killed and histopathology was analysed by haematoxylin and eosin staining of tissue sections. The competitive index was calculated as the \log_{10} of the strain ratio at a particular time point corrected by the inoculum ratio. Avirulent mutants were detected using a 'colony Western blot assay' for SipC. Genome re-sequencing was done at more than 50× coverage with an Illumina HiSeq2000 by GATC Biotech (NCBI Bioproject accession number PRJNA184189). Mutations were identified by alignment with the reference strain SL1344 (GenBank accession FQ312003.1) using the company pipeline. Statistical analysis was performed using individual mice as units of replication. Each data point in Figs 2–4 depicts data from a single mouse.

Full Methods and any associated references are available in the online version of the paper.

Received 4 August 2012; accepted 16 January 2013.

- West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms. *Nature Rev. Microbiol.* **4**, 597–607 (2006).
- Stecher, B. *et al.* *Salmonella enterica* serovar typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol.* **5**, e244 (2007).
- Kaiser, P., Diard, M., Stecher, B. & Hardt, W. D. The streptomycin mouse model for *Salmonella* diarrhea: functional analysis of the microbiota, the pathogen's virulence factors, and the host's mucosal immune response. *Immunol. Rev.* **245**, 56–83 (2012).

- Ackermann, M. *et al.* Self-destructive cooperation mediated by phenotypic noise. *Nature* **454**, 987–990 (2008).
- Sturm, A. *et al.* The cost of virulence: retarded growth of *Salmonella* Typhimurium cells expressing type III secretion system 1. *PLoS Pathog.* **7**, e1002143 (2011).
- Casadevall, A. & Pirofski, L. A. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infect. Immun.* **67**, 3703–3713 (1999).
- van Baalen, M. & Sabelis, M. W. The dynamics of multiple infection and the evolution of virulence. *Am. Nat.* **146**, 881–910 (1995).
- Lysenko, E. S., Lijek, R. S., Brown, S. P. & Weiser, J. N. Within-host competition drives selection for the capsule virulence determinant of *Streptococcus pneumoniae*. *Curr. Biol.* **20**, 1222–1226 (2010).
- Brown, S. P. Cooperation and conflict in host-manipulating parasites. *Proc. R. Soc. Lond. B* **266**, 1899–1904 (1999).
- Brown, S. P., Hochberg, M. E. & Grenfell, B. T. Does multiple infection select for raised virulence? *Trends Microbiol.* **10**, 401–405 (2002).
- Buckling, A. & Brockhurst, M. A. Kin selection and the evolution of virulence. *Heredity* **100**, 484–488 (2008).
- Harrison, F., Browning, L. E., Vos, M. & Buckling, A. Cooperation and virulence in acute *Pseudomonas aeruginosa* infections. *BMC Biol.* **4**, 21 (2006).
- Köhler, T., Buckling, A. & van Delden, C. Cooperation and virulence of clinical *Pseudomonas aeruginosa* populations. *Proc. Natl Acad. Sci. USA* **106**, 6339–6344 (2009).
- Rumbaugh, K. P. *et al.* Quorum sensing and the social evolution of bacterial virulence. *Curr. Biol.* **19**, 341–345 (2009).
- Raymond, B., West, S. A., Griffin, A. S. & Bonsall, M. B. The dynamics of cooperative bacterial virulence in the field. *Science* **337**, 85–88 (2012).
- Brown, S. P., Le Chat, L. & Taddei, F. Evolution of virulence: triggering host inflammation allows invading pathogens to exclude competitors. *Ecol. Lett.* **11**, 44–51 (2008).
- Endt, K. *et al.* The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal *Salmonella* diarrhea. *PLoS Pathog.* **6**, e1001097 (2010).
- Suar, M. *et al.* Accelerated type III secretion system 2-dependent enteropathogenesis by a *Salmonella enterica* serovar enteritidis PT4/6 strain. *Infect. Immun.* **77**, 3569–3577 (2009).
- Golubeva, Y. A., Sadik, A. Y., Ellermeier, J. R. & Schlauch, J. M. Integrating global regulatory input into the *Salmonella* pathogenicity island 1 type III secretion system. *Genetics* **190**, 79–90 (2012).
- Baxter, M. A., Fahlen, T. F., Wilson, R. L. & Jones, B. D. HilE interacts with HilD and negatively regulates *hilA* transcription and expression of the *Salmonella enterica* serovar Typhimurium invasive phenotype. *Infect. Immun.* **71**, 1295–1305 (2003).
- Cremer, J., Melbinger, A. & Frey, E. Growth dynamics and the evolution of cooperation in microbial populations. *Sci. Rep.* **2**, 281 (2012).
- Lawley, T. D. *et al.* Host transmission of *Salmonella enterica* serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. *Infect. Immun.* **76**, 403–416 (2008).
- Bush, K. *et al.* Tackling antibiotic resistance. *Nature Rev. Microbiol.* **9**, 894–896 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Slack, R. Kümmerli, M. Sellin and L. Robert for comments on the manuscript, G. Paul for input on the modelling, and Hardt laboratory members for discussions. M.D. was supported in part by the Fondation pour la Recherche Médicale. M.A., R.R.R. and W.-D.H. were supported by grants from the Swiss National Science Foundation.

Author Contributions M.D., V.G. and R.R.R. conceived and analysed the mathematical simulations. M.D., V.G. and R.R.R. wrote the theoretical part of the paper. M.D., W.-D.H., L.M. (Supplementary Figs 14 and 20), M.N.P.R.-E (Supplementary Fig. 8) and M.A. designed the experiments and analysed the data. M.D., W.-D.H. and M.A. wrote the paper. M.D. and L.M. (Supplementary Figs 14 and 20) performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.R.R. (roland.regoes@env.ethz.ch), M.A. (martin.ackermann@env.ethz.ch) or W.-D.H. (wolf-dietrich.hardt@micro.biol.ethz.ch).

METHODS

Model implementation. We used the R language of statistical computing²⁴ and the package deSolve²⁵ to implement the model.

Bacterial strains and plasmids. All *S. typhimurium* strains derived from *S. typhimurium* SL1344 SB300, a naturally streptomycin-resistant wild-type strain of *S. enterica* serovar Typhimurium²⁶ and the isogenic mutant SB161 (ref. 27) (*S. typhimurium invG*) was provided by J. E. Galan. The deletions of *hilA*, *hilE*, *hilD*, *tsr*, SPI-1 and *marT* (strains M3100, M3102, M3101, M3128, M3125 and M3020, respectively) were constructed as follows: the targeted gene was replaced by the chloramphenicol resistance (*cat*) cassette from pKD3 using λ /red homologous recombination as described²⁸. The construction was then P22-transduced into a clean SB300 background and *cat* was removed with the flipase encoded on pCP20 (ref. 28). The deletion of *stfC* was obtained by the transduction of the *stfC::aphT* construction from the mutant library provided by M. McClelland²⁹ in SB300, and the *aphT* cassette was removed by the flipase (strain M3127). For long-term infection experiments, each strain carried a deletion of *ssaV* that encodes an essential component of the *S. typhimurium* type III secretion system 2 (T2)³⁰. As the T2 is involved in the survival of the bacteria in phagocytic cells and is required for maximal and lethal colonization of systemic sites³¹, the use of attenuated mutants of *S. typhimurium* ensured host survival during the 10-day experiments, and allowed us to focus exclusively on the T1-dependent intestinal colitis. The *ssaV::cat* and *ssaV::aphT* constructions were transduced into the respective strains to disrupt the functioning of T2, and were used as antibiotic-resistance markers conferring resistance to 6 $\mu\text{g ml}^{-1}$ chloramphenicol or 50 $\mu\text{g ml}^{-1}$ kanamycin, respectively. The *Salmonella enterica* serovar Enteritidis (*S. enteritidis*) *ssaV::cat* strain (*S. enteritidis**, M1512)³² was used as a negative control during FACS experiments. The plasmids *philD* and *philA* derived from pBAD-Myc/HisC³³ (Invitrogen), itself a pBAD24 derivative³⁴. They conferred resistance to 100 $\mu\text{g ml}^{-1}$ ampicillin and allowed the controlled expression of genes of interest via the *araBAD* promoter (induced by 0.01% arabinose). The reporter plasmid pM972 carrying the fusion *psicA::gfp*^{35,36} (a reporter plasmid expressing green fluorescent protein (GFP) under control of the *sicA* promoter) and the transduced chromosomal cassette *pprgH::gfp*³⁷ (GFP under control of the *prgH* promoter) were used to monitor *ttss-1* expression. The suicide plasmid pAM34 (ref. 38) was used to evaluate the growth defect associated with *ttss-1* expression *in vivo*. This plasmid can only be replicated in the presence of isopropyl- β -D-thiogalactoside (IPTG). In broth culture, it was maintained by selection for ampicillin resistance (100 $\mu\text{g ml}^{-1}$) in the presence of IPTG (1 mM).

FACS analysis of *ttss-1* expression. GFP emission was analysed at 530 nm (FACSCalibur 4-colour, Becton Dickinson). Bacteria were identified by side scatter. Data were analysed with FlowJo software (Tree Star).

FACS analysis of T1[−] and T1⁺ cells from caecal content. Streptomycin-pretreated C57BL/6 mice were infected with either *S. typhimurium***PprgH::gfp* pAM34 or *S. enteritidis**. After 20 h, the animals were killed and the caecal contents were collected and homogenized (Qiagen Tissuizer) in 1 ml of ice-cold PBS containing 1% BSA. Tissues and food debris were eliminated by filtration through a 70- μm cell strainer (BD Falcon). Cells were then incubated for 1 h on ice with the *Salmonella* O antiserum factor 5 (Difco), pelleted by centrifugation (1 min at 8,500g at 4 °C) and resuspended in PBS plus 1% BSA. Afterwards, cells were incubated for 1 h on ice with the Cy5-labelled goat-anti-rabbit IgG Fabs (Jackson) and washed in ice-cold PBS. FACS was performed using a BD Aria III cell sorter. For gating, we used the background signals obtained with caecal material from an *S. enteritidis**-infected mouse, as *S. enteritidis** infects as efficiently as *S. typhimurium*^{32,39}, but is not recognized by the anti-lipopolysaccharide antiserum and it does not encode *gfp*. Immediately after sorting, cells were diluted and plated on MacConkey agar plates containing appropriate antibiotics and IPTG (1 mM).

Fluorescent *in situ* hybridization. C57BL/6 mice were infected for 24 h with *S. typhimurium***PprgH::gfp* or with *S. typhimurium**, killed, and the caecal contents were then collected and homogenized (Qiagen Tissuizer) in 1 ml of an ice-cold solution of 4% sucrose, 4% paraformaldehyde in PBS (Sigma). Fixation was performed for 1 h on ice and another hour at room temperature. Cells were washed with PBS and kept at 4 °C. Fixed cells were placed onto eight wells of a Teflon-coated slide (Tekdon), which had been treated with gelatin (Sigma) and air-dried before hybridization as described⁴⁰. In brief, cells were dehydrated in successive baths of ethanol (50%, 80% and 100%), 3 min each, air dried, covered by the hybridization buffer (90 μM NaCl, 20 μM Tris-HCl, pH 8, 35% formamide, 0.01% SDS (Sigma)) containing 1 μM of the Sal3-Cy3 probe (6 ng μl^{-1}) (5'-AATCACTTCACCTACGTG-3') (Microsynth) and incubated for 90 min at 46 °C in a sealed hybridization chamber. Cells were then washed in the washing buffer (70 μM NaCl, 20 μM Tris-HCl, pH 8, 10 μM EDTA (Sigma)) for 15 min at 48 °C, and dipped into ice-cold distilled water before air drying and mounting using VectaShield (Vector Laboratories). Images were taken using an Axioscope 2

microscope (Zeiss) equipped with a Hamamatsu ORCA ER digital camera (oil objective Plan-Apochromat $\times 63$, numerical aperture 1.4). The GFP signal was recorded using a FITC filter block (excitation bandpass 450–490 nm, emission bandpass 515–565 nm), and the Cy3 signal with a TRITC filter block (excitation bandpass 546–612 nm, emission bandpass 575–640 nm). A control Cy5 signal was recorded using the far red filter block (excitation bandpass 575–625 nm, emission bandpass 660–710 nm). This signal served to exclude autofluorescent particles. Image analysis was performed using Fiji⁴¹. Mean Cy3 and GFP intensity values were automatically gathered by defining particles of interest using a threshold on the Cy3 channel.

Within-host experimental evolution, competitions and short-term single infections. Eight-to-twelve-week-old C57BL/6 or 129 SvEv mice were treated with 25 mg streptomycin per os 24 h before inoculation as described⁴² to alleviate the colonization resistance. For infection, bacteria were grown in 0.3 M NaCl LB medium supplemented with the appropriate antibiotic for 12 h at 37 °C, and subcultured for 4 h in the same medium without antibiotic (T1-expression-inducing conditions). Around 5×10^7 bacteria, washed in PBS, were inoculated by gavage. Population sizes of *Salmonella* were monitored by plating serial dilutions of fresh faeces daily on MacConkey agar plates supplemented with appropriate antibiotics. Faeces were kept in PBS and frozen for further detection of lipocalin-2. At 10 d.p.i. mice were killed. Parts of the caecum were embedded in OCT for haematoxylin and eosin staining and pathoscore⁴². The competitions between two strains were performed under the same conditions. The competitive index was calculated as the \log_{10} of the ratio of competitor population sizes at a particular time point corrected by the initial ratio in the inoculum. To trigger colitis artificially and independently of virulent *S. typhimurium* when required, a dextran sulphate sodium (DSS) treatment was administered to the mice as described previously⁴³. In brief, sterile-filtered drinking water was supplemented with 5% DSS (molecular mass 36,000–50,000 g mol^{−1}, MP Biomedicals) and provided to the mice *ad libitum* from 1 to 6 d.p.i.

Colony blots. To detect the expression of SipC in isolated clones of *S. typhimurium*, we slightly modified the procedure recommended by Qiagen (<http://www.qiagen.com>). In brief, up to 350 isolated colonies from inoculum or faeces of infected mice were replica-transferred onto nitrocellulose membranes. Membranes were placed onto LB agar supplemented with 50 $\mu\text{g ml}^{-1}$ streptomycin (bacteria facing to the top). After overnight growth at 37 °C, colonies were lysed by passing membranes on successive discs of Whatman paper soaked with the following buffers: 10% SDS for 10 min, denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 5 min, two 5-min washes in neutralization solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.4) and 15 min on 2 \times SSC (3 M NaCl, 0.3 M sodium citrate, pH 7). Membranes were then washed twice in TBS (10 mM Tris-HCl, 150 mM NaCl, pH 7.5) and the remaining debris of the colonies was gently scraped with Whatman paper. TBS plus 3% BSA was used to block the membrane for 1 h at room temperature before incubation overnight at 4 °C in a moisture chamber with anti-SipC serum. After washing steps in TBS-T (20 mM TrisHCl, 500 mM NaCl, 0.05% Tween 20 and 0.2% Triton X-100, pH 7.5) and TBS, membranes were incubated with horseradish peroxidase (HRP)-coupled anti-rabbit secondary antibody (Santa Cruz Biotechnology) for 1 h at room temperature, and washed twice in TBS-T buffer. The blots were developed using 4-chloro-1-naphthol as a chromogenic substrate for HRP in the presence of 3% hydrogen peroxide in TBS. All reagents were from Sigma unless stated otherwise.

Genome re-sequencing. All ancestors and evolved clones were sequenced to more than 50 \times coverage using the Illumina HiSeq2000 platform by the GATC Biotech Company. Mutations (that is, single nucleotide polymorphisms and small insertion/deletions) were identified on the basis of the sequence of the reference strain SL1344 (chromosome sequence and plasmids sequences were respectively obtained from GenBank (FQ312003.1) and the Genoscope database (<https://www.genoscope.cns.fr>) using the company pipeline. Only mutations found in at least 95% of the reads were considered as true positives.

Targeted re-sequencing. The *hilD* gene was re-sequenced by Microsynth using the Sanger technique on PCR products generated with the following primers: *hilD*-contr.: 5'-GCGATGTCTGTCGTCTCGATAGC-3', and *hilD*-contr.2: 5'-ATAGCCTCCCATCTGATAG-3'.

Western blot. SipC was detected using an anti-SipC antiserum in a lysate of 1×10^6 bacteria grown under inducing conditions for *ttss-1* expression.

Lipocalin-2 dosage. Lipocalin-2 was detected in homogenized faeces by ELISA according to instructions provided by the DuoSet ELISA kit manufacturer (R&D Systems).

Statistical analyses. Mann-Whitney U tests and Wilcoxon matched pairs signed rank tests were performed using GraphPad Prism version 5.00 for Windows (<http://www.graphpad.com>).

Ethics statement. All experiments were approved (licences 201/2007 and 223/2010 Kantonales Veterinäramt Zürich) and performed according to local guidelines (TschV, Zurich) and the Swiss animal protection law (TschG).

24. R Development Core Team. A Language and Environment for Statistical Computing <http://www.R-project.org> (R Foundation for Statistical Computing, 2010).
25. Soetaert, K., Petzoldt, T. & Setzer, R. W. Solving differential equations in R: Package deSolve. *J. Stat. Softw.* **33**, 1–25 (2010).
26. Hoiseth, S. K. & Stocker, B. A. Aromatic-dependent *Salmonella* typhimurium are non-virulent and effective as live vaccines. *Nature* **291**, 238–239 (1981).
27. Kaniga, K., Bossio, J. C. & Galan, J. E. The *Salmonella* typhimurium invasion genes *invF* and *invG* encode homologues of the AraC and PulD family of proteins. *Mol. Microbiol.* **13**, 555–568 (1994).
28. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
29. Santiviago, C. A. *et al.* Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog.* **5**, e1000477 (2009).
30. Browne, S. H., Hasegawa, P., Okamoto, S., Fierer, J. & Guiney, D. G. Identification of *Salmonella* SPI-2 secretion system components required for SpvB-mediated cytotoxicity in macrophages and virulence in mice. *FEMS Immunol. Med. Microbiol.* **52**, 194–201 (2008).
31. Hensel, M. *et al.* Genes encoding putative effector proteins of the type III secretion system of *Salmonella* pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. *Mol. Microbiol.* **30**, 163–174 (1998).
32. Suar, M. *et al.* Accelerated type III secretion system 2-dependent enteropathogenesis by a *Salmonella enterica* serovar enteritidis PT4/6 strain. *Infect. Immun.* **77**, 3569–3577 (2009).
33. Schechter, L. M. & Lee, C. A. AraC/XylS family members, HilC and HilD, directly bind and derepress the *Salmonella* typhimurium *hilA* promoter. *Mol. Microbiol.* **40**, 1289–1299 (2001).
34. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
35. Sturm, A. *et al.* The cost of virulence: retarded growth of *Salmonella* Typhimurium cells expressing type III secretion system 1. *PLoS Pathog.* **7**, e1002143 (2011).
36. Stecher, B. *et al.* Flagella and chemotaxis are required for efficient induction of *Salmonella enterica* serovar Typhimurium colitis in streptomycin-pretreated mice. *Infect. Immun.* **72**, 4138–4150 (2004).
37. Hautefort, I., Proenca, M. J. & Hinton, J. C. Single-copy green fluorescent protein gene fusions allow accurate measurement of *Salmonella* gene expression in vitro and during infection of mammalian cells. *Appl. Environ. Microbiol.* **69**, 7480–7491 (2003).
38. Gil, D. & Bouche, J. P. ColE1-type vectors with fully repressible replication. *Gene* **105**, 17–22 (1991).
39. Endt, K. *et al.* The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal *Salmonella* diarrhoea. *PLoS Pathog.* **6**, e1001097 (2010).
40. Nordentoft, S., Christensen, H. & Wegener, H. C. Evaluation of a fluorescence-labelled oligonucleotide probe targeting 23S rRNA for *in situ* detection of *Salmonella* serovars in paraffin-embedded tissue sections and their rapid identification in bacterial smears. *J. Clin. Microbiol.* **35**, 2642–2648 (1997).
41. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).
42. Barthel, M. *et al.* Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. *Infect. Immun.* **71**, 2839–2858 (2003).
43. Slack, E. *et al.* Innate and adaptive immunity cooperate flexibly to maintain host-microbiota mutualism. *Science* **325**, 617–620 (2009).

Abundant SAR11 viruses in the ocean

Yanlin Zhao^{1*}, Ben Temperton^{1*}, J. Cameron Thrash¹, Michael S. Schwalbach², Kevin L. Vergin¹, Zachary C. Landry¹, Mark Ellisman³, Tom Deerinck³, Matthew B. Sullivan⁴ & Stephen J. Giovannoni¹

Several reports proposed that the extraordinary dominance of the SAR11 bacterial clade in ocean ecosystems could be a consequence of unusual mechanisms of resistance to bacteriophage infection, including ‘cryptic escape’ through reduced cell size¹ and/or K-strategist defence specialism². Alternatively, the evolution of high surface-to-volume ratios coupled with minimal genomes containing high-affinity transporters enables unusually efficient metabolism for oxidizing dissolved organic matter in the world’s oceans that could support vast population sizes despite phage susceptibility. These ideas are important for understanding plankton ecology because they emphasize the potentially important role of top-down mechanisms in predation, thus determining the size of SAR11 populations and their concomitant role in biogeochemical cycling. Here we report the isolation of diverse SAR11 viruses belonging to two virus families in culture, for which we propose the name ‘pelagiphage’, after their host. Notably, the pelagiphage genomes were highly represented in marine viral metagenomes, demonstrating their importance in nature. One of the new phages, HTVC010P, represents a new podovirus subfamily more abundant than any seen previously, in all data sets tested, and may represent one of the most abundant virus subfamilies in the biosphere. This discovery disproves the theory that SAR11 cells are immune to viral predation and is consistent with the interpretation that the success of this highly abundant microbial clade is the result of successfully evolved adaptation to resource competition.

Four pelagiphages were isolated from seawater samples taken on the Oregon coast and at Bermuda Hydrostation S, using a host culture of axenic ‘*Candidatus Pelagibacter ubique*’ HTCC1062. SAR11 isolates do not produce lawns on agar plates, therefore dilution-to-extinction³ was used to purify the viruses (see Supplementary Methods). Infections of axenic *Pelagibacter* cell cultures by the isolated viruses caused sharp declines in host numbers and produced phage particles (Supplementary Fig. 1). Pelagiphage HTVC011P, HTVC019P and HTVC010P all morphologically belong to the short-tailed Podoviridae family (Fig. 1 a–c). The general biological and genomic features of these four pelagiphages are summarized in Table 1. HTVC011P and HTVC019P have similar morphologies, with capsid sizes of approximately 55 nm, whereas HTVC010P has a smaller capsid (approximately 50 nm in diameter). HTVC008M has the typical morphology of the Myoviridae family with an isometric head and contractile tail structure (Fig. 1d). Transmission electron microscopy also captured a SAR11 host cell filled with HTVC011P immediately before lysis (Fig. 1e). The latent periods of these four pelagiphages ranged from 16 to 24 h (Table 1 and Supplementary Fig. 2). Their burst sizes varied from 9 to 49 (Table 1). The genomes of all four pelagiphages were completely sequenced, assembled and annotated, yielding genome sequences of 34,892–147,284 base pairs (bp) (Table 1). Phylogenetic analysis of the pelagiphages confirmed morphological classification of HTVC011P and HTVC019P as podoviruses, HTVC008M as a myovirus, and HTVC010P as a divergent podovirus (Fig. 2). Genome annotation revealed that both HTVC011P and HTVC019P are members of the

Autographivirinae subfamily within the Podoviridae family⁴ with striking conservation of synteny (Supplementary Fig. 3). HTVC011P and HTVC019P are divergent from marine cyanophage subgroups and other well-defined Autographivirinae subgroups, showing closer relationships to some Autographivirinae prophages (Fig. 2a). HTVC008M is most closely related to T4-like myoviruses (Supplementary Fig. 4), but again represents a divergent lineage distinct from the marine cyanophage clusters and other T4-like subgroups⁵ (Fig. 2b). The genome of HTVC010P contains limited homology to some unclassified podoviruses (Supplementary Fig. 5). Single-gene trees of the tail tube B protein and the head–tail connector protein placed HTVC010P near unclassified podoviruses distantly related to Autographivirinae subfamily podoviruses (Fig. 2c, d).

We investigated the relative abundances of pelagiphages in a recently published collection of quantitative marine viral metagenomes from the Pacific Ocean⁶. The viral metagenomes were prepared using chemical flocculation of viral particles to yield sufficient DNA material followed by linker-amplified libraries modified for 454 pyrosequencing to avoid stochastic amplification biases traditionally associated with multiple displacement amplification (MDA)⁷. Sequences of pelagiphage origin were characterized by reciprocal-best-BLAST (RBB) (see Supplementary Methods) in 27 metagenomes from two multi-depth coastal-to-open ocean transects and one from Scripps pier, San Diego, California, USA⁶, and their abundances were compared to marine cyanophages. Across all viral metagenomic samples, a total of 183,657 reads (~4.7% of total reads) were successfully assigned to one of the 30 viral genomes used in this study (Supplementary

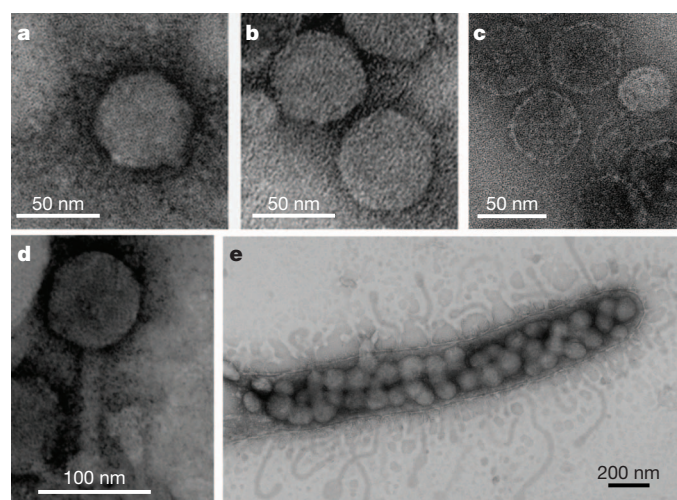


Figure 1 | Transmission electron microscopy images of isolated pelagiphages. a, Pelagipodovirus HTVC011P. b, Pelagipodovirus HTVC019P. c, Pelagipodovirus HTVC010P. d, Pelagimyovirus HTVC008M. e, Host cell of ‘*Candidatus P. ubique*’ HTCC1062 infected with HTVC011P immediately before lysis.

¹Department of Microbiology, Oregon State University, Corvallis, Oregon 97331, USA. ²Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Wisconsin 53706, USA. ³National Center for Microscopy and Imaging Research, University of California, San Diego, California 92093, USA. ⁴Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA.

*These authors contributed equally to this work.

Table 1 | General features of pelagiphages analysed in this study

| Phage | Source water | Taxon | Capsid size (mean \pm s.d., nm) | Latent period (h) | Burst size (mean \pm s.d.) | Genome size (bp) | Number of ORFs | G+C % | Accession number |
|----------|----------------|-------------|--------------------------------------|----------------------|---------------------------------|---------------------|-------------------|-------|------------------|
| HTVC011P | Oregon NH10 | Podoviridae | 55 \pm 2 | 19–22 | 49 \pm 5 | 39,921 | 45 | 32.0 | KC465900 |
| HTVC019P | Oregon NH10 | Podoviridae | 55 \pm 1 | 19–22 | 37 \pm 5 | 42,102 | 59 | 34.0 | KC465901 |
| HTVC010P | Hydrostation S | Podoviridae | 50 \pm 3 | 22–24 | 42 \pm 7 | 34,892 | 64 | 29.7 | KC465898 |
| HTVC008M | Hydrostation S | Myoviridae | 84 \pm 3 | 16–19 | 9 \pm 2 | 147,284 | 200 | 33.5 | KC465899 |

ORFs, open reading frames.

Table 3). Results shown in Fig. 3 revealed that the pelagiphages were among the most abundant phages in all metagenomic data sets across longitudinal and depth gradients. When normalized for database size and total amino-acid length of the predicted open reading frames for each genome, 58.7% of successfully assigned reads were assigned to pelagiphages (with 38.8% assigned to HTVC010P), followed by 21.6%, 12.4% and 6% for phages infecting *Synechococcus*, *Prochlorococcus* and *Roseobacter*, respectively. Pelagiphages were more abundant in the photic zone than in the aphotic zone (Supplementary Fig. 6). The abundances of T4-like and Autographivirinae pelagiphages were both comparable to the cyanophages from the same phage subfamily. HTVC010P represents a new subfamily of podovirus with high abundance in Pacific Ocean data sets. The abundance of HTVC010P-like

podoviruses was compared to the cumulative abundance of members of T4-like myoviruses and Autographivirinae podoviruses (Fig. 2). HTVC010P-like podoviruses were 2.5 \times more abundant than all T4-like cyanophages combined and 4.9 \times more abundant than all Autographivirinae cyanophages combined, and thus represent a highly abundant phage group, which includes members that infect SAR11. Fragment recruitment of metagenomic reads to the concatenated predicted open reading frames of the four pelagiphage genomes showed significant recruitment to genes encoding structural proteins and polymerases (Supplementary Fig. 7). RBB analysis of the viral genomes against four MDA-amplified marine viral metagenomes from the Sargasso Sea, the Arctic Ocean, the Gulf of Mexico and British Columbia coastal waters⁸ showed significant representation of the

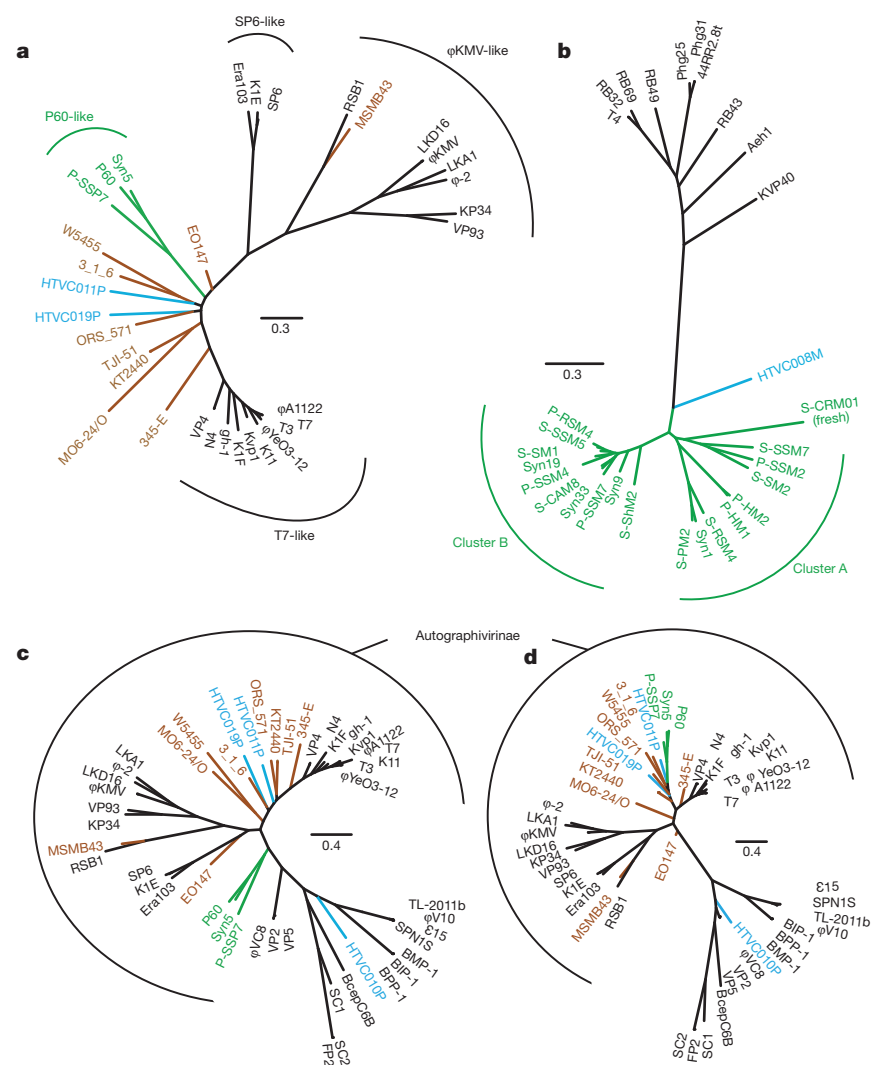


Figure 2 | Phylogenetic analysis of the pelagiphages. Bootstrap values have been omitted for clarity, but all trees are available at <http://giannonilab.science.oregonstate.edu/publications>. Blue, pelagiphage; green, cyanophage; brown, prophage; black, other. Scale bars indicate changes per position. **a**, Maximum-likelihood (ML) tree of HTVC011P, HTVC019P and near

neighbours within the Autographivirinae subfamily using 16 concatenated proteins. **b**, ML tree of HTVC008M and near neighbours using 35 concatenated proteins. **c**, ML tree of the tail tube protein B for the Podoviridae. **d**, ML tree of the head-tail connector protein for the Podoviridae. See Supplementary Tables 1 and 2 for information about phage genomes.

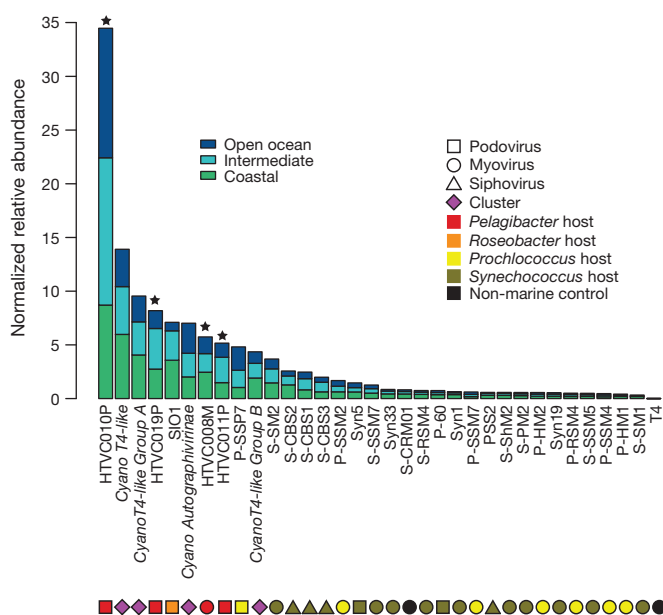


Figure 3 | Abundance of pelagiphage relative to representative cyanophages and roseophage SIO1. Linker-amplified marine viral metagenomic data sets were prepared from two Pacific transects and one surface sample taken at Scripps pier, San Diego, California, USA⁶. Relative abundances were calculated by reciprocal best-BLAST (see Supplementary Methods) and normalized by total number of sequences in each database category (open ocean, intermediate, and coastal) and by total number of amino acids in the viral query sequences. Stars denote pelagiphage genomes identified in this study. To highlight the relative abundance of HTVC010P compared to known cyanophage subfamilies, the abundance of subfamily members identified in Fig. 2 were summed and are represented as clusters (italics). Non-marine negative control sequences (Enterobacteria phage T4 and freshwater *Synechococcus* phage S-CRM01) were included in the analysis.

four pelagiphage genomes beyond the Pacific Ocean. RBB analysis against a viral metagenome from an Antarctic lake⁹ revealed a small number of HTVC010P-like fragments (Supplementary Fig. 8).

The viruses we report here, particularly HTVC010P, are highly represented in oceans. Quantification of viral fragments by RBB analysis is sensitive to the number of closely related genomes within the database (Supplementary Fig. 9). Therefore, the high number of HTVC010P-like sequences recruited from the database could be in part due to its genomic novelty, including hits comprised of both HTVC010P-like pelagiphages and HTVC010P-like phages infecting other unknown hosts. We propose that if closely related representatives of HTVC010P were relatively infrequent, compared to the other viruses we studied, the average per cent amino acid identities and bitscores of sets of reads recruited by RBB would be lower. However, in three separate analyses using all genes, the top five most highly recruited genes and the top recruited gene for each phage revealed average per cent amino acid identity and bitscores for HTVC010P RBB hits were not lower than most other phage (Supplementary Fig. 10), supporting the interpretation that the high abundances measured for HTVC010P are not a consequence of their novelty causing 'greedy' recruitment. Furthermore, analyses of legacy data sets with the HTVC010P genome sequence revealed very similar sequences highly represented in a shotgun fosmid library from the deep chlorophyll maximum in the Mediterranean¹⁰. Eight out of 197 (~4%) fosmid sequences contained lengthy regions of similarity ($74.4 \pm 1.2\%$ nucleotide identity, $n = 65$, 95% confidence interval), some spanning multiple coding and non-coding sequences (Supplementary Fig. 11). Given the high natural variability of phage genomes, this striking result shows that phages very closely related to those we isolated can be recovered from other oceans by random cloning procedures.

Studies of SAR11 genomes have identified a number of putative genes of phage origin and phage defence mechanisms, suggesting that the evolution of SAR11 has occurred in an environment where phage predation is common. A 48-kbp hypervariable region (HVR2) bounded by RNA and tRNA genes has been reported previously¹¹. Comparative genomics of seven SAR11 genomes, spanning the proposed order 'Pelagibacterales'¹², has revealed that HVR2 is conserved across distantly related SAR11 strains, and is rich in paralogues and genes involved in synthesis of cell surface proteins, including various transferases, isomerases, O-antigen polymerases and pilins¹². Thus, hypervariable regions in SAR11 genomes may be involved in viral resistance by modification of cell-surface proteins to restrict phage adsorption¹³. Similarly, a recent investigation using co-cultures of *Prochlorococcus* strains and cyanophages showed that genes within regions of hypervariable genomic islands accumulate nonsynonymous mutations under phage predation¹⁴. One SAR11 strain was reported to possess a phosphorothioation system that modifies the DNA backbone¹⁵, which we speculate may provide immunity to some phages. Viral susceptibility and processes involved in immunity and defence to phage infection across the SAR11 clade are likely to be diverse and remain unexplored, but are now open to investigation as model phage–host systems of demonstrated ecological relevance are available for study.

The 'Kill-the-Winner' hypothesis is a model for understanding the impact of top-down control of community dynamics in microbial systems^{16,17}. The essence of this hypothesis is that rising population densities and cell metabolic activity expose planktonic cells to more efficient viral predation by increasing host–phage encounter rate and phage propagation capability. Evidence supporting this hypothesis has been demonstrated both in cultures¹⁸ and in natural populations¹⁹. Consequently, the global dominance of SAR11 in marine communities has led to speculative theories that SAR11 avoids phage predation either by having a very small cell size¹ or by growing so slowly as to make infection inefficient^{2,20}. However, whereas SAR11 is one of the smallest bacteria in the oceans²¹, the surface area and volume of SAR11 cells are comparable to those of *Prochlorococcus* spp., for which multiple phages have been isolated and whose ecology is clearly affected by phages^{22–25}. Furthermore, the contribution of SAR11 to bacterial heterotrophic production in the open ocean is greater than their relative abundance, particularly when phosphate is limiting²⁶, suggesting that SAR11 does not avoid viral predation by growing slowly.

We propose that the dominance of SAR11 in marine bacterial communities is likely to be a result of their superior competitiveness for nutrient uptake and that, once achieved, high host abundance provides certain protections against population decimation by viral predation. High population densities increase encounter rates with infective phage particles, but they also increase the probability that SAR11 cells will encounter conspecific cells or their DNA, and have the opportunity for genetic recombination by conjugation or transformation. Recombination enables genomic elements to propagate in populations more rapidly than is possible by clonal replication and thus potentially offers an advantage to the spread of immunity. Indeed, previous work on non-recombining species demonstrated high susceptibility to 'Kill-the-Winner' selection²⁷. Although a detailed mechanism for recombination remains to be elucidated, *pil* genes involved in conjugation and DNA uptake are conserved in genomes of the dominant SAR11 strains HTCC7211 and HTCC1062 (ref. 12). Furthermore, recombination rates in SAR11 populations from a single sample of surface water were reported previously to be among the highest on record^{28,29}. Therefore, we propose that high recombination rates in SAR11 populations, particularly in the HVRs, allow for rapid adaptation to novel phage phenotypes. This mechanism for co-evolution and co-existence of pelagiphage and SAR11 is archetypal of the 'Red Queen Hypothesis'³⁰ and addresses the paradox of SAR11 as both abundant and susceptible to phage, in apparent conflict with the 'Kill-the-Winner' hypothesis.

METHODS SUMMARY

Four pelagiphages infecting marine SAR11 strain 'Candidatus Pelagibacter ubique' HTCC1062 were isolated from Oregon coast seawater samples (HTVC011P and HTVC019P) and Bermuda Hydrostation S seawater samples (HTVC010P and HTVC008M). Phages were purified by dilution-to-extinction method³. Phage growth curves and host mortality were determined in infected cultures of 'Candidatus P. ubique' HTCC1062 and phage morphologies were visualized by transmission electron microscopy. Pelagiphage genomes were sequenced with a combination of pyrosequencing (HTVC008M, HTVC010P, HTVC019P) and shotgun cloning and primer walking (HTVC011P) before assembly and annotation. Phylogenetic analysis of the pelagiphages was performed with maximum-likelihood trees using both single structural genes and concatenated proteins from conserved predicted open reading frames. Predicted open reading frames from pelagiphage genomes were also used to estimate relative abundances of pelagiphages with respect to known important marine cyanophages and roseophage SIO1 in quantitative Pacific Ocean viral metagenomes⁶ using a reciprocal best-BLAST approach¹¹ with modifications (see Supplementary Methods). A significant presence of pelagiphages in legacy viral metagenomes^{8,9} as well as a fosmid metagenome from the Mediterranean deep chlorophyll maximum¹⁰ was also observed using a similar approach. Conservation of genes encoding structural proteins within metagenomes was confirmed in all pelagiphages by fragment recruitment analysis. A complete description of the materials and methods is provided in the Supplementary Information.

Received 12 October 2012; accepted 18 January 2013.

Published online 13 February 2013.

1. Yoosseph, S. *et al.* Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**, 60–66 (2010).
2. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nature Rev. Microbiol.* **5**, 801–812 (2007).
3. Nagasaki, K. & Bratbak, G. In *Manual of Aquatic Viral Ecology Limnol. Oceanogr.* 92–101 (2010).
4. Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W. & Kropinski, A. M. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. *Res. Microbiol.* **159**, 406–414 (2008).
5. Ignacio-Espinoza, J. C. & Sullivan, M. B. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ. Microbiol.* **14**, 2113–2126 (2012).
6. Hurwitz, B. L. & Sullivan, M. B. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* (in the press) (2013).
7. Duhaime, M. B., Deng, L., Poulos, B. T. & Sullivan, M. B. Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ. Microbiol.* **14**, 2526–2537 (2012).
8. Angly, F. E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4**, e368 (2006).
9. López-Bueno, A. *et al.* High diversity of the viral community from an Antarctic lake. *Science* **326**, 858–861 (2009).
10. Ghai, R. *et al.* Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J.* **4**, 1154–1166 (2010).
11. Wilhelm, L. J., Tripp, H. J., Givan, S. A., Smith, D. P. & Giovannoni, S. J. Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol. Direct* **2**, 27 (2007).
12. Grote, J. *et al.* Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* **3**, <http://dx.doi.org/10.1128/mBio.00252-12> (2012).
13. Rodríguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nature Rev. Microbiol.* **7**, 828–836 (2009).
14. Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. & Lindell, D. Genomic island variability facilitates *Prochlorococcus*–virus coexistence. *Nature* **474**, 604–608 (2011).
15. Wang, L. *et al.* DNA phosphorothioation is widespread and quantized in bacterial genomes. *Proc. Natl Acad. Sci. USA* **108**, 2963–2968 (2011).
16. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27 (1997).
17. Chow, C.-E. T. & Fuhrman, J. A. Seasonality and monthly dynamics of marine myovirus communities. *Environ. Microbiol.* **14**, 2171–2183 (2012).
18. Marston, M. F. *et al.* Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc. Natl Acad. Sci. USA* **109**, 4544–4549 (2012).
19. Fuhrman, J. A. & Schwalbach, M. Viral influence on aquatic bacterial communities. *Biol. Bull.* **204**, 192–195 (2003).
20. Parsons, R. J., Breitbart, M., Lomas, M. W. & Carlson, C. A. Ocean time-series reveals recurring seasonal patterns of viroplankton dynamics in the northwestern Sargasso Sea. *ISME J.* **6**, 273–284 (2012).
21. Rappé, M. S., Connon, S. A., Vergin, K. L. & Giovannoni, S. J. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**, 630–633 (2002).
22. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
23. Sullivan, M. B., Coleman, M., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3**, e144 (2005).
24. Sullivan, M. B. *et al.* The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ. Microbiol.* **11**, 2935–2951 (2009).
25. Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* **12**, 3035–3056 (2010).
26. Malmstrom, R. R., Cottrell, M. T., Elifantz, H. & Kirchman, D. L. Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Appl. Environ. Microbiol.* **71**, 2979–2986 (2005).
27. Kuniin, V. *et al.* A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res.* **18**, 293–297 (2008).
28. Vergin, K. L. *et al.* High intraspecific recombination rate in a native population of *Candidatus Pelagibacter ubique* (SAR11). *Environ. Microbiol.* **9**, 2430–2440 (2007).
29. Vos, M. & Didelot, X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J.* **3**, 199–208 (2009).
30. Van Valen, L. A new evolutionary law. *Evol. Theory* **1**, 1–30 (1973).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Tucson Marine Phage Lab and B. Hurwitz for early access to the Pacific Ocean virome datasets and J. Yan for her assistance in isolating HTVC011P. This work was supported by an investigator award to S.J.G. from the Gordon and Betty Moore Foundation Marine Microbiology Initiative.

Author Contributions Y.Z. isolated phages and performed genome sequencing and annotation; B.T. designed and implemented metagenomic bioinformatic analyses and prepared the manuscript; J.C.T. performed phylogenetic and metagenomic analyses; M.S.S. began the project and isolated the first virus, HTVC011P; Y.Z., M.E. and T.D. performed transmission electron microscopy; K.L.V. sequenced the HTVC011P genome; Z.C.L. assembled and annotated the viral genomes. M.B.S. and S.J.G. gave technical support and conceptual advice. S.J.G. assisted in writing the manuscript.

Author Information Viral genome sequences and annotations have been deposited at GenBank/EMBL/DBJ under accession codes: KC465898 (HTVC10P), KC465899 (HTVC008M), KC465900 (HTVC011P) and KC465901 (HTVC019P). Scripts and data used to generate the figures in this manuscript are available at <http://giovannonilab.science.oregonstate.edu/publications>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.J.G. (steve.giovannoni@oregonstate.edu).

T-helper-1-cell cytokines drive cancer into senescence

Heidi Braumüller^{1*}, Thomas Wieder^{1*}, Ellen Brenner¹, Sonja Aßmann¹, Matthias Hahn¹, Mohammed Alkhaled², Karin Schilbach², Frank Essmann³, Manfred Kneilling¹, Christoph Griessinger^{1,4}, Felicia Ranta⁵, Susanne Ullrich⁵, Ralph Mocikat⁶, Kilian Braungart¹, Tarun Mehra¹, Birgit Fehrenbacher¹, Julia Berdel¹, Heike Niessner¹, Friedegund Meier¹, Maries van den Broek⁷, Hans-Ulrich Häring⁵, Rupert Handgretinger^{2,8}, Leticia Quintanilla-Martinez⁹, Falko Fend^{8,9}, Marina Pesic¹⁰, Jürgen Bauer¹, Lars Zender¹⁰, Martin Schaller¹, Klaus Schulze-Osthoff^{3,8} & Martin Röcken^{1,8}

Cancer control by adaptive immunity involves a number of defined death^{1–8} and clearance^{9–11} mechanisms. However, efficient inhibition of exponential cancer growth by T cells and interferon- γ (IFN- γ) requires additional undefined mechanisms that arrest cancer cell proliferation^{1–5,12,13}. Here we show that the combined action of the T-helper-1-cell cytokines IFN- γ and tumour necrosis factor (TNF) directly induces permanent growth arrest in cancers. To safely separate senescence induced by tumour immunity from oncogene-induced senescence^{9–11,14–17}, we used a mouse model in which the Simian virus 40 large T antigen (Tag) expressed under the control of the rat insulin promoter creates tumours by attenuating p53- and Rb-mediated cell cycle control^{18,19}. When combined, IFN- γ and TNF drive Tag-expressing cancers into senescence by inducing permanent growth arrest in G1/G0, activation of p16INK4a (also known as CDKN2A), and downstream Rb hypophosphorylation at serine 795. This cytokine-induced senescence strictly requires STAT1 and TNFR1 (also known as TNFRSF1A) signalling in addition to p16INK4a. *In vivo*, Tag-specific T-helper 1 cells permanently arrest Tag-expressing cancers by inducing IFN- γ - and TNFR1-dependent senescence. Conversely, *Tnfr1*^{−/−} Tag-expressing cancers resist cytokine-induced senescence and grow aggressively, even in TNFR1-expressing hosts. Finally, as IFN- γ and TNF induce senescence in numerous murine and human cancers, this may be a general mechanism for arresting cancer progression.

Recent studies from targeted cancer immunotherapies show that adaptive immunity can efficiently control human cancer^{20–24}. Many cancer immunotherapies do not cause cytotoxic cancer cell elimination, but instead arrest cancer growth or induce slow cancer regression^{21,22}, despite the fact that immunotherapeutic strategies generally focus on CD8⁺ cytotoxic T lymphocytes (CTL) or natural killer cells^{1–8,23,24}. Moreover, where studied, growth arrest and cancer regression correlate with tumour-specific, IFN- γ -producing CD4⁺ T-helper 1 (T_H1) cells rather than CTLs^{20–23}. In addition, profiling of patients in clinical cancer trials shows a critical role for IFN- γ and TNF in cancer control^{25,26}.

Similarly, efficient immune control of murine cancers resulting from aberrant cell cycle control, oncogene expression and chemical or viral transformation strictly requires IFN- γ ^{3–7,12,27}. In consequence, IFN- γ - and TNF-producing T_H1 cells specific for the tumour antigen Tag (Tag-T_H1 cells) restrain Tag-induced islet cancers in mice expressing Tag

under the control of the rat insulin promoter (RIP-Tag2 mice; in which Tag2 indicates a mouse strain that carries five copies of the transgene) in all pancreatic islet cells^{12,18}. T_H1 immunity doubles the lifespan of mice through strictly IFN- γ - and TNFR1-dependent mechanisms¹², without causing detectable signs of cytotoxicity, tumour cell necrosis or apoptosis¹². This was surprising, as Tag expression causes invasive β -cell cancers (β -cancers) in 2% of the islets by incomplete Rb suppression and p53 silencing^{18,19}.

Only CD4⁺ T_H1 cells that produce IFN- γ and TNF and are specific for Tag peptide induce IFN- γ - and TNFR1-dependent arrest of RIP-Tag2 tumours (Supplementary Fig. 1a–c)¹². This arrest occurs also in the absence of notable T-cell infiltration (Supplementary Figs 1c and 2)¹² and is independent of both CTLs¹² and enhanced apoptosis, as determined by TdT-mediated dUTP nick end labelling (TUNEL) assay¹² and caspase 3 staining, respectively. Moreover, IFN- γ -exposed β -cancer cells failed to express major histocompatibility complex (MHC) class II *in vivo* as well as *in vitro*, a prerequisite for cancer killing by T_H1 cells. In consequence, Tag-T_H1 cells failed to kill IFN- γ -stimulated β -cancer cells *in vitro* (Supplementary Fig. 3). In addition, as Tag-T_H1 cells enhance, rather than attenuate, the growth of *Tnfr1*^{−/−} β -cancers¹² (Supplementary Fig. 1), TNFR1-independent killing—for example, by perforin-granzyme B-mediated lysis—could not account for substantial control of β -cancers by Tag-T_H1 cells. Instead, Tag-T_H1 cells form follicle-like structures around the islets, where they interact with antigen-presenting cells¹². Ki67 staining confirmed that Tag-T_H1 cells arrested proliferation of β -cancers *in vivo* (Supplementary Fig. 4a). Freshly isolated β -cancer cells from Tag-T_H1-cell-treated RIP-Tag2 mice failed to proliferate *in vitro*, whereas β -cancer cells from sham-treated mice strongly incorporated ³H-thymidine (Supplementary Fig. 4b). This suggests that T_H1 cytokines induce growth arrest *in vivo*, despite normal Tag expression (Supplementary Fig. 5).

To assess directly whether IFN- γ and TNF induce senescence in β -cancers, we cultured freshly isolated β -cancer cells from sham-treated mice treated with either control medium, or with IFN- γ and TNF. In cell cycle analysis, untreated β -cancer cells were $\geq 25\%$ in S phase and 40% in G1/G0, explaining their rapid proliferation. When combined, IFN- γ and TNF arrested most β -cancer cells in G1/G0 within 3 days (Fig. 1a). They reduced S-phase cells to 3% and increased the G1/S ratio 20-fold (Fig. 1b) without increasing the apoptotic sub-G1 fraction (Fig. 1a). As IFN- γ and TNF arrested the β -cancer cells in G1/G0, a state characterizing cellular senescence, we wanted to know

¹Department of Dermatology, Eberhard Karls University, Liebermeister Strasse 25, 72076 Tübingen, Germany. ²Department of General Pediatrics, Oncology/Hematology, Eberhard Karls University, Hoppe-Seyler-Strasse 1, 72076 Tübingen, Germany. ³Interfaculty Institute for Biochemistry, Eberhard Karls University, Hoppe-Seyler-Strasse 4, 72076 Tübingen, Germany. ⁴Laboratory for Preclinical Imaging and Imaging Technology of the Werner Siemens-Foundation, Department for Preclinical Imaging and Radiopharmacy, Eberhard Karls University, Röntgenweg 13, 72076 Tübingen, Germany. ⁵Department of Internal Medicine IV, Endocrinology, Diabetology and Clinical Chemistry, Eberhard Karls University, Otfried-Müller-Strasse 10, 72076 Tübingen, Germany. ⁶Institut für Molekulare Immunologie, Helmholtz-Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt, Marchionini Strasse 25, 81377 München, Germany. ⁷Clinic of Oncology, University Hospital Zurich, Raemistrasse 100, 8091 Zürich, Switzerland. ⁸Comprehensive Cancer Center Tübingen, German Cancer Consortium (DKTK), Herrenberger Strasse 23, D-72070 Tübingen, Germany. ⁹Department of Pathology, Eberhard Karls University, Liebermeister Strasse 8, 72076 Tübingen, Germany. ¹⁰Division of Molecular Oncology of Solid Tumors, Department of Internal Medicine I, Eberhard Karls University, Otfried-Müller-Strasse 10, 72076 Tübingen, Germany.

*These authors contributed equally to this work.

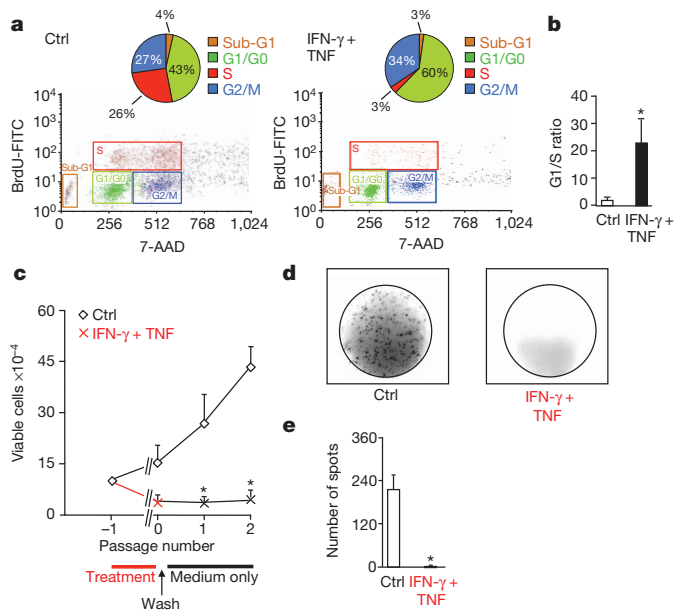


Figure 1 | Combined, the T_H1 cytokines IFN- γ and TNF induce stable growth arrest of Tag-driven β -cancer cells *in vitro*. **a**, **b**, Cell cycle analysis (**a**) and mean G1/S-phase ratio (**b**) of β -cancer cells cultured either in the presence or absence of IFN- γ and TNF. Ctrl, control; FITC, fluorescein isothiocyanate. 7-AAD, 7-Aminoactinomycin. **c**–**e**, Cell numbers (**c**), BrdU incorporation (**d**) and mean numbers of BrdU-positive spots (**e**) of β -cancer cells treated for 5 days with medium only, or with IFN- γ and TNF. Thereafter, cells were washed and then cultured with medium in the absence of cytokines for another two passages. After the second passage 3,000 viable cells were seeded in 96-well plates and cell proliferation was analysed by BrdU staining. * $P < 0.05$. Error bars are mean \pm s.e.m. $n = 3$ –6 (**b**, **c**, **e**).

whether IFN- γ and TNF caused senescence-defining permanent growth arrest.

Freshly isolated β -cancer cells proliferated rapidly in medium. When cultured in the presence of IFN- γ and TNF, β -cancer cells were fully growth arrested (Fig. 1c). To determine whether the growth-arrested β -cancer cells were really senescent, we washed the cells on day 5 and cultured them for another 2 weeks with medium only. Although untreated β -cancer cells continued to expand, β -cancer cells that had been exposed for 5 days to the combined action of IFN- γ and TNF were truly senescent, as they remained fully growth arrested (Fig. 1c). Even 2 weeks after withdrawal of IFN- γ and TNF, the β -cancer cells failed to incorporate 5-bromo-2-deoxyuridine (BrdU), whereas untreated controls strongly incorporated BrdU (Fig. 1d, e and Supplementary Fig. 6).

IFN- γ and TNF also induced characteristic senescence-associated epigenetic and lysosomal changes, such as nuclear recruitment of phosphorylated heterochromatin protein 1 γ (pHP1 γ) into senescence-associated heterochromatin foci or senescence-associated β -galactosidase (SA- β -gal) activity. Time-course studies revealed that within 3 days, IFN- γ and TNF induced the early senescence marker pHP1 γ in 75% (Supplementary Fig. 7a, b) and SA- β -gal in 50% of the β -cancer cells (Fig. 2a, b). However, induction of stable growth arrest and the late senescence marker SA- β -gal in 80% of the cells needed ≥ 4 days of incubation with both IFN- γ and TNF (Fig. 2c). Double-staining with synaptophysin, an islet cell marker, confirmed SA- β -gal expression by β -cancer cells (Supplementary Fig. 8). Combined IFN- γ and TNF treatment established the senescence-defining permanent growth arrest in β -cancers cells, whereas neither IFN- γ nor TNF alone was sufficient to induce full growth arrest (Supplementary Fig. 6), although early signs of senescence, such as pHP1 γ recruitment to senescence-associated heterochromatin foci (Supplementary Fig. 7a, b) and SA- β -gal positivity in 40% of the cells, were observed (Fig. 2c).

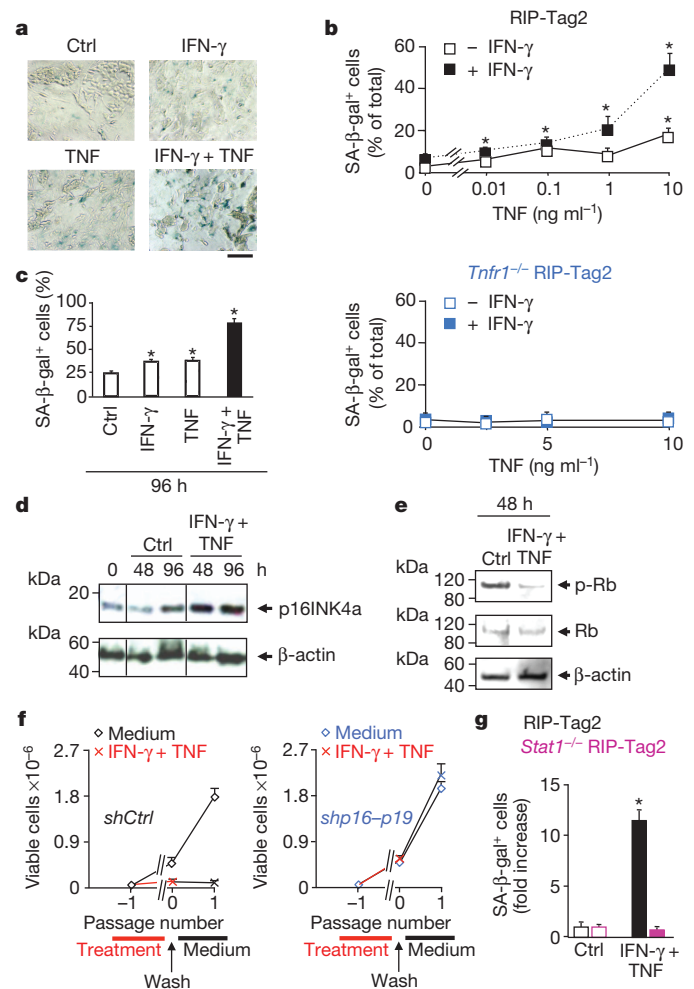


Figure 2 | STAT1- and TNFR1-dependent stabilization of the p16INK4a-Rb senescence pathway in β -cancer cells by the combined action of IFN- γ and TNF *in vitro*. **a**, SA- β -gal activity of β -cancer cells after 72 h of treatment with medium, IFN- γ , TNF, or IFN- γ and TNF. Scale bar, 100 μ m. **b**, Concentration-dependent induction of SA- β -gal $^{+}$ cells by TNF, either in the presence or absence of IFN- γ , within 72 h in β -cancer cells from RIP-Tag2 or *Tnfr1* $^{-/-}$ RIP-Tag2 mice. **c**, Induction of SA- β -gal activity by IFN- γ , TNF, or IFN- γ and TNF within 96 h. **d**, Detection of p16INK4a or β -actin by western blotting in β -cancer cells treated with medium or with IFN- γ and TNF. **e**, Detection of phosphorylated Rb (p-Rb), total Rb or β -actin by western blotting in β -cancer cells treated with medium or IFN- γ and TNF. **f**, β -cancer cells were transduced with *shCtrl* or *shp16-p19* murine stem cell virus (MSCV) vectors, and then treated for 5 days with medium only, or with IFN- γ and TNF. Thereafter, cells were washed and cultured with medium in the absence of cytokines. Mean cell numbers of three cultures are shown. **g**, SA- β -gal activity of β -cancer cells isolated from either RIP-Tag2 or *Stat1* $^{-/-}$ RIP-Tag2 mice after 72 h of treatment with medium or with IFN- γ and TNF. * $P < 0.05$. Error bars are mean \pm s.e.m. $n = 3$ –9 (**b**, **c**, **f**, **g**).

As combined stimulation of islets or islet tumours with IFN- γ and TNF strongly induces JUNB²⁸, the combined IFN- γ STAT1 and TNFR1 signalling may activate the JUNB downstream target p16INK4a and thus stabilize the p16INK4a-Rb senescence pathway in Tag-expressing β -cancers. Indeed, IFN- γ and TNF strongly induced p16INK4a in subconfluent cultures within 48 h (Fig. 2d), whereas p16INK4a remained weak in medium-treated subconfluent β -cancer cells (Fig. 2d). This induction of p16INK4a also caused sustained and severe hypophosphorylation of Rb at Ser 795 (Fig. 2e). Corresponding to the conjoint IFN- γ - and TNF-induced senescence through stabilization of the p16INK4a-Rb pathway in β -cancer cells, short hairpin RNA (shRNA)-mediated knockdown of *p16INK4a* and *p19* (hereafter termed *p16-p19*, in which *p19* refers to an alternate reading frame of

the *Ink4a/Arf* (*Cdkn2a*) locus) fully abrogated senescence induction by IFN- γ and TNF (Fig. 2f and Supplementary Fig. 9). In line with this, STAT1- or TNFR1-deficient β -cancer cells fully resisted senescence induction by IFN- γ and TNF, showing that hypophosphorylation of Rb strictly requires the combined action of the STAT1 and TNFR1 signalling pathways (Fig. 2b, g).

IFN- γ and TNF induced senescence in breast cancers from polyomavirus middle T antigen (PyVMT)-transgenic mice (Supplementary Fig. 10), in 3 out of 6 mouse tumour lines (Supplementary Table 1), in human A204 rhabdomyosarcoma cells (Supplementary Fig. 11), in 6 out of 11 IFN-receptor- and TNF-receptor-expressing human cancers from the National Cancer Institute's NCI-60 panel (Supplementary Table 2), and in freshly isolated primary melanoma or sarcoma cells (Supplementary Table 3). Interestingly, primary melanomas frequently show areas of inflammation and regression. Such tumour areas harboured senescent melanoma cells strongly expressing p16INK4a but not the proliferation marker Ki67, whereas p16INK4a-deficient melanoma cells strongly expressed Ki67 even in the presence of an inflammatory infiltrate (Supplementary Fig. 12). Thus, cytokine-induced senescence was not restricted to Tag-expressing β -cancers, but seems to be of broad relevance for tumour immunotherapy and may cause tumour dormancy in sporadic human melanomas.

Combined IFN- γ and TNFR1 signalling is also necessary to arrest β -cancers *in vivo*¹² (Supplementary Fig. 1) and a senescence-like phenotype was observed in spontaneously regressing melanomas. Thus, the *in vitro* and *in vivo* data suggest that T_H1 immunity arrested cancer progression through IFN- γ - and TNF-induced senescence *in vivo*. To test this hypothesis, we first quantified senescence-associated chromatin changes by counting cancer cells positive for pHP1 γ , trimethylation of histone 3 lysine 9 (H3K9me3), or apoptosis-associated active caspase 3 by immunohistochemistry in either sham-treated mice or mice treated with Tag-T_H1 cells. T_H1 immunity significantly increased pHP1 γ - and H3K9me3-positive nuclei in β -cancers, but not caspase-3-positive cells (Fig. 3a, b).

To address whether T_H1 immunity also activated the p16INK4a-Rb pathway *in vivo*, we double-stained sections for p16INK4a and the proliferation marker Ki67. In sham-treated mice, the β -cancer cells were $\geq 30\%$ Ki67 positive and only 5% p16INK4a positive. T_H1 immunity diminished Ki67⁺ cells to 3%, and increased the fraction of p16INK4a⁺ cells (nuclear and cytoplasmic) to $\geq 20\%$ (Fig. 3c, d and Supplementary Fig. 13).

Senescence induction by T_H1 immunity *in vivo* also strictly required TNFR1 signalling. *Tnfr1*^{-/-} β -cancers failed to increase pHP1 γ , H3K9me3 or active caspase 3 when treated with Tag-T_H1 cells (Supplementary Fig. 14a, b). In addition, whether isolated from sham- or Tag-T_H1-treated mice, *Tnfr1*^{-/-} β -cancer cells expressed Ki67 but not p16INK4a (Fig. 3c, d), even though Tag-T_H1 cells migrate into pancreata of *Tnfr1*^{-/-} RIP-Tag2 mice¹². As T_H1 immunity severely impaired β -cancer growth *in vivo*¹², these data strongly suggest that the combined IFN- γ and TNFR1 signalling also drove cancers into senescence *in vivo*.

The most stringent criterion defining senescence is stable and permanent growth arrest in the absence of T_H1 immunity^{9,15,16}. We therefore cultured freshly isolated β -cancer cells from 12-week-old mice with medium only. Cells from sham-treated RIP-Tag2 mice first suffered a critical loss and then re-initiated proliferation, yielding $10\text{--}20 \times 10^5$ β -cancer cells per pancreas after three passages (Fig. 4a). By contrast, β -cancer cells from RIP-Tag2 mice treated with Tag-T_H1 cells were truly senescent, as they failed to proliferate over three passages and stably yielded only $0.5\text{--}2.0 \times 10^5$ cells per pancreas (Fig. 4a). The T_H1-induced growth arrest remained stable when cultures were extended for six passages (Supplementary Fig. 15). At 12 weeks, β -cancers from *Tnfr1*^{-/-} RIP-Tag2 mice yielded similar cell counts as those from RIP-Tag2 mice (Fig. 4a). *In vitro*, *Tnfr1*^{-/-} β -cancer cells grew exponentially, yielding 100×10^5 cells within three passages even when derived from Tag-T_H1-treated mice (Fig. 4a).

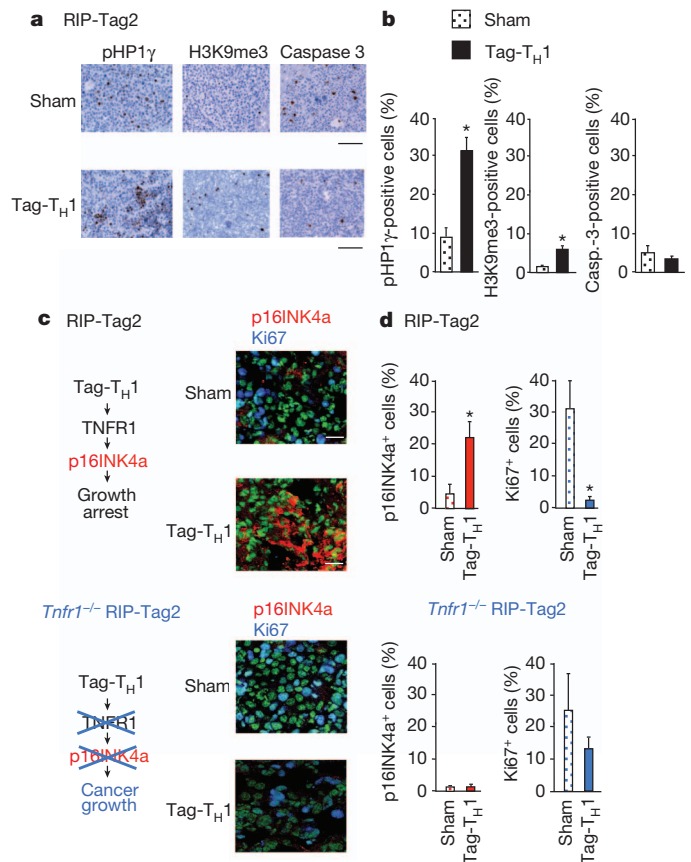


Figure 3 | TNFR1-dependent induction of growth arrest and senescence in β -cancer cells by T_H1 immunity *in vivo*. **a, b**, Analysis of the senescence markers pHP1 γ and H3K9me3, or the apoptosis marker active caspase 3, by immunohistochemistry (**a**), and percentage of pHP1 γ -, H3K9me3- or active caspase-3-positive cells in β -cancers from RIP-Tag2 mice that were either sham- or Tag-T_H1-cell-treated (**b**). Scale bars, 100 μ m. **c, d**, Double-staining for the senescence marker p16INK4a (red) and the proliferation marker Ki67 (blue) (**c**), and percentage of p16INK4a- and Ki67-positive cells in cancers from RIP-Tag2 or *Tnfr1*^{-/-} RIP-Tag2 mice that were sham- or Tag-T_H1-cell-treated (**d**). Nuclei are depicted in green. Scale bars, 25 μ m. **P* < 0.05 from sham-treated control. Error bars are mean \pm s.e.m. *n* = 5–6 (**b, d**).

To determine whether senescence was also maintained *in vivo*, we re-implanted the four different β -cancer cell lines after the third passage under the skin of non-obese diabetic-severe combined immunodeficient (NOD-SCID) *Il2rg*^{-/-} mice, as growth at ectopic sites characterizes cancers². Within 7 weeks, as few as 1.2×10^5 β -cancer cells from sham-treated RIP-Tag2 mice significantly decreased blood glucose (Fig. 4b) and increased the serum insulin levels (Fig. 4c), demonstrating the metastatic potential of β -cancer cells from sham-treated mice. Senescent β -cancers from RIP-Tag2 mice treated with Tag-T_H1 cells also remained growth arrested *in vivo*, as blood glucose remained stable in all transplanted mice throughout the 7 weeks (Fig. 4b). Minute adenomas in some NOD-SCID *Il2rg*^{-/-} mice and marginally increased serum insulin (Fig. 4c) showed that the transplanted cells survived but remained growth arrested for ≥ 10 weeks of *in vitro* and *in vivo* culture. Again, senescence-resistant *Tnfr1*^{-/-} RIP-Tag2 β -cancer cells grew rapidly after transplantation and as little as 1.2×10^5 *Tnfr1*^{-/-} RIP-Tag2 cancer cells rapidly decreased blood glucose (Fig. 4b), irrespective of whether they were derived from sham-treated or Tag-T_H1-treated mice. Transplanting 60% of total β -cancer cells from sham-treated RIP-Tag2 mice generated tumours within 7 weeks, whereas β -cancer cells from Tag-T_H1-cell-treated mice failed to grow (Fig. 4d). Only 10% of β -cancer cells from sham- or Tag-T_H1-treated *Tnfr1*^{-/-} RIP-Tag2 mice generated large tumours within the same timeframe (Fig. 4d).

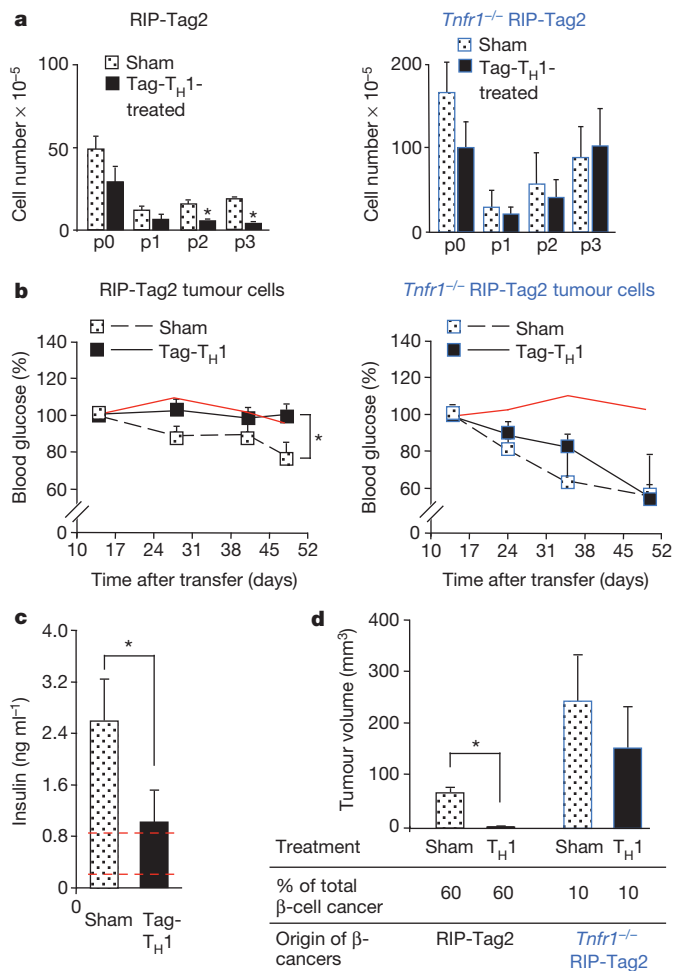


Figure 4 | T_H1 immunity induces TNFR1-dependent, permanent growth arrest of β-cancer cells that remains stable for ≥10 weeks, even after transfer into immune-deficient NOD-SCID *Il2rg*^{-/-} mice. **a**, Cell numbers of β-cancer cells isolated from 12-week-old RIP-Tag2 or *Tnfr1*^{-/-} RIP-Tag2 mice that were sham- or Tag-T_H1-cell-treated. The number of cell passages (p) is indicated, and the data are presented as number of living β-cancer cells per mouse. **b**, Blood glucose levels in NOD-SCID *Il2rg*^{-/-} mice after transfer of 1.2×10^5 β-cancer cells isolated from sham- or Tag-T_H1-cell-treated RIP-Tag2 or *Tnfr1*^{-/-} RIP-Tag2 mice. Red lines indicate blood glucose of untreated NOD-SCID *Il2rg*^{-/-} mice. **c**, Insulin levels in NOD-SCID *Il2rg*^{-/-} mice after transfer of 1.2×10^5 β-cancer cells isolated from sham- or Tag-T_H1-cell-treated RIP-Tag2 mice. Dashed red lines indicate normal range of insulin in healthy mice. **d**, Top, tumour volumes in NOD-SCID *Il2rg*^{-/-} mice after transfer of β-cancer cells isolated from sham- or Tag-T_H1-cell-treated mice. Bottom, origin and percentage of total β-cancer cells per mouse injected. **P* < 0.05 from sham-treated control. Error bars are mean ± s.e.m. *n* = 3–6 (**a–d**).

Oncogenes, DNA damage and chemotherapeutics induce senescence that, in human cells, is reinforced by the senescence-associated secretome^{14–17,29,30}. Uncovering that adaptive T_H1 immunity directly restrains cancer proliferation through IFN-γ- and TNF-induced cancer cell senescence provides a long-sought-for direct mechanism explaining the anti-proliferative effects of T_H1 immunity on cancers^{2,3}. T_H1 immunity can induce antiangiogenic chemokines¹², even in cancer cells (Supplementary Fig. 16a, b), and permanent growth arrest leading to cancer cell senescence. Although both effects protect against cancer^{12,18}, *p16–p19* knockdown experiments revealed that the arrest of cancer growth strictly needed T_H1-induced senescence. shRNA-mediated knockdown of *p16–p19* affected neither the IFN-γ- and TNF-induced mRNA expression nor protein production of antiangiogenic chemokines, but it fully abrogated senescence. Following the *p16–p19* knockdown, β-cancer cells grew exponentially in the

presence of IFN-γ, TNF and the antiangiogenic chemokines CXCL9 and CXCL10 (Fig. 2f and Supplementary Fig. 16c, d). Importantly, this is also valid *in vivo*. Despite the absence of antiangiogenic signals, senescent cancers remained growth arrested when transferred into NOD-SCID *Il2rg*^{-/-} mice (Fig. 4b, c). As the T_H1-induced cancer cell senescence thus explains the therapeutic efficiency of tumour-specific T_H1 immunity in early cervical cancer²¹ and disseminated melanoma²⁰, T_H1-cytokine-induced senescence may be of broad relevance for cancer control, also in humans under therapeutic conditions.

METHODS SUMMARY

RIP-Tag2 and *Tnfr1*^{-/-} RIP-Tag2 mice were either sham-treated or treated with Tag2-specific T_H1 cells starting at week 6. Cancer cells were isolated by consecutive collagenase-trypsin digestion from tumour tissues of RIP-Tag2, *Stat1*^{-/-} RIP-Tag2, *Tnfr1*^{-/-} RIP-Tag2, or PyVmT-transgenic mice. Isolated β-cancer cells were identified by immunofluorescence using anti synaptophysin (early β-cell marker) antibodies. Proliferation *in vivo* and *in vitro* of tumour cells was measured by BrdU incorporation, Ki67 staining, ³H-thymidine incorporation or cell cycle analysis. Senescence was assessed by SA-β-gal staining, immunofluorescence, immunohistochemistry or western blotting using anti-pHP1γ, anti-p16INK4a, anti-Rb, anti-phospho-Rb, or anti-H3K9me3 antibodies, or by *in vitro* and *in vivo* growth assays. Apoptosis was determined by cell cycle analysis, or immunohistochemistry with an anti-active-caspase-3 antibody. *p16INK4a* was knocked down in β-cancer cells by the use of *shp16–p19* MSCV vectors. For transfer experiments, β-cancer cells from sham- or T_H1-cell-treated mice (either RIP-Tag2 or *Tnfr1*^{-/-} RIP-Tag2) were injected subcutaneously into NOD-SCID *Il2rg*^{-/-} mice, and tumour growth was monitored with a caliper and by measuring blood glucose and insulin levels.

Full Methods and any associated references are available in the online version of the paper.

Received 11 April; accepted 6 December 2012.

Published online 3 February 2013.

1. Finn, O. J. Cancer immunology. *N. Engl. J. Med.* **358**, 2704–2715 (2008).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
3. Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565–1570 (2011).
4. Koebel, C. M. *et al.* Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903–907 (2007).
5. van den Broek, M. E. *et al.* Decreased tumor surveillance in perforin-deficient mice. *J. Exp. Med.* **184**, 1781–1790 (1996).
6. Willimsky, G. & Blankenstein, T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. *Nature* **437**, 141–146 (2005).
7. Mocikat, R. *et al.* Natural killer cells activated by MHC class I^{low} targets prime dendritic cells to induce protective CD8 T cell responses. *Immunity* **19**, 561–569 (2003).
8. Hung, K. *et al.* The central role of CD4⁺ T cells in the antitumor immune response. *J. Exp. Med.* **188**, 2357–2368 (1998).
9. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
10. Rakhra, K. *et al.* CD4⁺ T cells contribute to the remodeling of the microenvironment required for sustained tumor regression upon oncogene inactivation. *Cancer Cell* **18**, 485–498 (2010).
11. Kang, T. W. *et al.* Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. *Nature* **479**, 547–551 (2011).
12. Müller-Hermelink, N. *et al.* TNFR1 signaling and IFN-γ signaling determine whether T cells induce tumor dormancy or promote multistage carcinogenesis. *Cancer Cell* **13**, 507–518 (2008).
13. Röcken, M. Early tumor dissemination, but late metastasis: insights into tumor dormancy. *J. Clin. Invest.* **120**, 1800–1803 (2010).
14. Braig, M. *et al.* Oncogene-induced senescence as an initial barrier in lymphoma development. *Nature* **436**, 660–665 (2005).
15. Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
16. Collado, M. & Serrano, M. Senescence in tumours: evidence from mice and humans. *Nature Rev. Cancer* **10**, 51–57 (2010).
17. Nardella, C., Clohessy, J. G., Alimonti, A. & Pandolfi, P. P. Pro-senescence therapy for cancer treatment. *Nature Rev. Cancer* **11**, 503–511 (2011).
18. Bergers, G., Javaherian, K., Lo, K. M., Folkman, J. & Hanahan, D. Effects of angiogenesis inhibitors on multistage carcinogenesis in mice. *Science* **284**, 808–812 (1999).
19. Casanovas, O., Hager, J. H., Chun, M. G. & Hanahan, D. Incomplete inhibition of the Rb tumor suppressor pathway in the context of inactivated p53 is sufficient for pancreatic islet tumorigenesis. *Oncogene* **24**, 6597–6604 (2005).

20. Hunder, N. N. *et al.* Treatment of metastatic melanoma with autologous CD4+ T cells against NY-ESO-1. *N. Engl. J. Med.* **358**, 2698–2703 (2008).
 21. Kenter, G. G. *et al.* Vaccination against HPV-16 oncoproteins for vulvar intraepithelial neoplasia. *N. Engl. J. Med.* **361**, 1838–1847 (2009).
 22. Hodi, F. S. *et al.* Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
 23. Schwartzentruber, D. J. *et al.* gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N. Engl. J. Med.* **364**, 2119–2127 (2011).
 24. Morgan, R. A. *et al.* Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science* **314**, 126–129 (2006).
 25. Canova, C. *et al.* Genetic associations of 115 polymorphisms with cancers of the upper aerodigestive tract across 10 European countries: the ARCADE project. *Cancer Res.* **69**, 2956–2965 (2009).
 26. Critchley-Thorne, R. J. *et al.* Impaired interferon signaling is a common immune defect in human cancer. *Proc. Natl Acad. Sci. USA* **106**, 9010–9015 (2009).
 27. Zhang, B., Karrison, T., Rowley, D. A. & Schreiber, H. IFN- γ - and TNF-dependent bystander eradication of antigen-loss variants in established mouse cancers. *J. Clin. Invest.* **118**, 1398–1404 (2008).
 28. Gurzov, E. N. *et al.* Pancreatic β -cells activate a JunB/ATF3-dependent survival pathway during inflammation. *Oncogene* **31**, 1723–1732 (2012).
 29. Kuilman, T. *et al.* Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. *Cell* **133**, 1019–1031 (2008).
 30. Acosta, J. C. *et al.* Chemokine signaling via the CXCR2 receptor reinforces senescence. *Cell* **133**, 1006–1018 (2008).
- T. Haug for technical support in the chromium release assay, R. Dummer for melanoma samples, W. Kempf for technical support in the cell cycle analysis and S. Lowe for the *p16–p19* shRNA concept. The technical assistance of S. Weidemann and M. Dierstein is gratefully acknowledged. This work is part of the doctoral thesis of E.B., S.A., M.H., K.B., J.Berdel and C.G., and was supported by the Sander Stiftung (2005.043.2 and 2005.043.3), the Deutsche Krebshilfe (No. 109037), the IZKF-Promotionskolleg ‘Molekulare Medizin’ 2010 (1886-0-0), 2011 (PK 2011-3) and 2012 (PK 2012-1), the Deutsche Forschungsgemeinschaft (SFB 685, SFB 773 and Wi 1279/3-1) and in part by the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DzD e.V.).

Author Contributions M.R. originally developed the concept, further elaborated on it, and designed the experiments together with H.B., T.W., R.M. and K.S.-O. H.B., T.W., S.A., M.K., C.G., F.E., M.H., K.B., T.M. and E.B. performed experiments and analysed the data. B.F. and M.S. established and carried out fluorescence microscopy. L.Q.-M. and F.F. performed light microscopy and advised on cell biology. M.A., K.S. and R.H. supervised and performed the β -cancer-cell-transfer experiments. L.Z. and M.P. generated *shp16–p19* MSCV vectors and supervised the knockdown experiments. H.N. and F.M. isolated primary melanoma cells, J.Berdel and J.Bauer collected human melanoma specimen and performed immunohistochemical analysis. F.R., S.U. and H.-U.H. isolated β -cancer cells and advised on β -cell physiology. H.B., T.W., M.v.d.B., K.S.-O. and M.R. interpreted the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R. (mrocken@med.uni-tuebingen.de).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank A. Knuth, H. G. Rammensee, K. Ghoreschi, J. Brück, G. Riethmüller, G. Stingl, T. Biedermann and A. Yazdi for discussions,

METHODS

Animals. C3HeB/FeJ (C3H) mice from The Jackson Laboratory, transgenic RIP-Tag2 mice³¹, double-transgenic *Tnfr1*^{−/−} RIP-Tag2^{12,32} and *Stat1*^{−/−} RIP-Tag2 mice (backcross of *Stat1*^{−/−} mice from Taconic over 12 generations to C3H), TCR2 mice³³ all on a C3H background, and PyVMT-transgenic mice on a C57BL/6 background³⁴, were bred under specific pathogen-free conditions. NOD-SCID *Il2rg*^{−/−} mice (NOD.Cg-Prkdc^{scid} *Il2rg*^{tm1Wjl/Sz})³⁵ were from The Jackson Laboratory. Animal experiments were approved by the local authorities (Regierungspräsidium Tübingen, Germany; reference numbers HT 2/03, HT2/07 and K1/07).

Isolation of primary human cancer cells. Two human rhabdomyosarcoma and four human melanoma cancer cell preparations were isolated from patients. The isolation of cancer cells from patient-derived material and histology of cancers were approved by the local ethics committees (Ethik-Kommission an der Medizinischen Fakultät der Eberhard-Karls-Universität und am Universitätsklinikum Tübingen, Germany, reference numbers 072/2011BO2, 105/2008BO1 and 017/2012BO2; and Ethik-Kommission (KEK), Gesundheitsdirektion Kanton Zürich, Switzerland, reference number Nr. 647). Informed consent was obtained from all patients.

Cell culture and single-cell analysis. Tag-specific T_H1 cells were isolated and generated from female TCR2 mice and characterized by flow cytometry¹².

Tumours were isolated from sham- or Tag-T_H1-cell-treated female RIP-Tag2 mice, sham- or Tag-T_H1-cell-treated female *Tnfr1*^{−/−} RIP-Tag2 mice, female *Stat1*^{−/−} RIP-Tag2 mice, or mammary-tumour-bearing PyVMT-transgenic mice by collagenase digestion (1 mg ml^{−1}, Serva) for 10 min at 37 °C, and then separated under a dissection microscope (Leica Microsystems). Tumour cells were obtained by incubation in 0.05% trypsin/EDTA solution (Invitrogen) at 37 °C for 10 min, and seeded on tissue culture plates. Adherent cells were cultured for 2–7 weeks in RPMI 1640 supplemented with 10% fetal calf serum, non-essential amino acids, antibiotics and 50 µM 2-mercaptoethanol at 37 °C and 5% CO₂. The murine cancer cell lines B16, LLC and CT26 EpCam³⁶ in addition to 11 human cancer cell lines from the NCI-60 panel³⁷, six primary human cancer cell preparations and human rhabdomyosarcoma cells (A204 cells) were grown in complete RPMI 1640 medium. If not otherwise stated, subconfluent cells were treated with 100 ng ml^{−1} mouse or human IFN-γ (R&D Systems), or 10 ng ml^{−1} mouse or human TNF (R&D Systems), or with a combination of mouse or human IFN-γ (50–100 ng ml^{−1}) and mouse or human TNF (0.1–10 ng ml^{−1}) for 2–6 days. β-cancer cells were identified by immunofluorescence using an anti-synaptophysin antibody (undiluted; Lifespan Biosciences).

Knockdown of p16INK4a. 5 × 10⁴ β-cancer cells were seeded in 6-well culture plates. After 72 h, cells were transfected with 2 ml cell culture supernatant containing *shCtrl* or *shp16-p19* MSCV vectors^{38,39} in the presence of 1 µg ml^{−1} polybrene (Sigma) for a total transduction time of 12 h. After 5 days, transduced cells were selected by treatment with 1 µg ml^{−1} puromycin (Sigma) for 72 h. The transduction rate was determined by counting green fluorescent protein-positive cells under a Zeiss Axiovert 200 microscope (Zeiss).

In vitro proliferation assays. After treatment, cancer cell proliferation was measured either by the [³H]-thymidine-incorporation assay⁴⁰, or by the BrdU-based Cell Proliferation ELISA and XTT-based Cell Proliferation Kit II according to the manufacturer's protocols (Roche Diagnostics). [³H]-thymidine incorporation was quantified using a MicroBeta TriLux counter (PerkinElmer) and colorimetric analyses were performed on a Multiskan EX microplate reader (Thermo Electron).

In vitro growth-arrest assays. The different cancer cells were seeded at a density of 1 × 10⁴ cells cm^{−2}. Next, the cells were treated with control medium or cytokines as described above for 4–5 days. After treatment, the cells were trypsinized and viable cells (trypan blue exclusion) were counted under a Zeiss Axiovert 25 microscope (Zeiss) using a Neubauer counting-chamber (Karl Hecht GmbH). The cells were seeded at 2 × 10⁴ cells cm^{−2} and grown in complete RPMI 1640 medium until the control cultures reached confluence. Then, the cells were trypsinized, counted and seeded again. After passage 1–2, 1,000–3,000 viable cells were seeded on MultiscreenTM HTS 96 well Filtration Plates (Millipore), and proliferation was measured by the BrdU-based Cell Proliferation ELISA (Roche Diagnostics) in combination with the VectorSG Substrate Kit for Peroxidase from Vector Laboratories to visualize BrdU-incorporating cell clusters. BrdU-positive spots were counted with an ELISPOT reader (Bioreader-3000; Bio-Sys). In addition, some cultures were stained with 4',6-diamidino-2-phenylindole (DAPI; Invitrogen) to visualize the nuclei of adherent cells.

Analysis of CXCL9 or CXCL10 in the supernatant of β-cancer cells. β-cancer cells were either treated with medium alone or with IFN-γ and TNF. After 24 h incubation, CXCL9 or CXCL10 levels were determined in the cell culture supernatants using the mouse CXCL9 or mouse CXCL10 ELISA kit from R&D Systems.

Treatment of mice with Tag-T_H1 cells. Before the first Tag-T_H1-cell-based therapy, all mice received 2-Gy total-body irradiation. Then, 1 × 10⁷ Tag-T_H1 cells in

0.9% NaCl solution (Tag-T_H1) or NaCl solution alone (sham) was injected intraperitoneally once per week starting at week 6 (ref. 12). After 6 weeks of Tag-T_H1 treatment, all mice received a second 2-Gy total-body irradiation¹². Blood glucose was measured using an Accu-Check sensor (Roche Diagnostics). If not otherwise stated, mice were euthanized at week 12.

Transfer of β-cancer cells into immune-deficient mice. β-cancer cells isolated from the various groups of mice were cultured for three passages. Then, 10–60% of the β-cancer cells were injected subcutaneously into immune-deficient NOD-SCID *Il2rg*^{−/−} mice. Tumour growth was monitored with a caliper, and blood glucose was measured using an Accu-Check sensor for up to 7 weeks. Serum insulin levels were determined using the rat/mouse insulin ELISA kit from Millipore.

Immunofluorescence and immunohistochemistry. The different cancer cells were grown on chamber slides (BD Biosciences). After treatment, the cells were fixed with acetone/methanol 1:1. The slides were washed with PBS buffer and 0.05% Tween 20 at room temperature (21–23 °C), blocked with serum-free DAKO-Block (DAKO), washed again, and then incubated with the following antibodies: anti-Ki67 (dilution 1:100; Abcam), anti-PCNA (dilution 1:100; Cell Signaling Technology), anti-pHP1γ (dilution 1:100; Abcam), anti-H3K9me3 (dilution 1:500; Millipore), anti-p16INK4a (dilution 1:100; Santa Cruz Biotechnology), anti-SV40 Tag (dilution 1:100; BD Biosciences), anti-MHCII (dilution 1:50; eBioscience) or anti-synaptophysin (undiluted; Lifespan Biosciences). After washing, the slides were incubated with anti-rabbit Alexa488 (Invitrogen), anti-rabbit-Cy3 (Dianova), anti-mouse Alexa555, or anti-mouse Alexa488 (both from Cell Signaling Technology), washed again and incubated with DAPI (Invitrogen). Finally, the slides were washed, mounted with fluorescence mounting medium (DAKO) and analysed using a Zeiss Axiovert 200 microscope (Zeiss) with the VisiView software (Visitron Systems).

Fresh-frozen cryostat sections of whole pancreata were stained as described⁴⁰. In brief, the sections were fixed with periodate-lysine-paraformaldehyde, blocked with donkey serum (dilution 1:20), and then incubated with rabbit anti-pHP1γ (dilution 1:80), mouse anti-PCNA (dilution 1:50), mouse anti-p16INK4a (dilution 1:50), or rabbit anti-Ki67, rat-anti-MHCII (dilution 1:50), or rabbit-anti-synaptophysin (undiluted) antibodies. After washing, the sections were incubated with Cy3-conjugated donkey anti-rabbit, donkey anti-mouse, or Cy5-conjugated donkey-anti-rabbit, donkey-anti-mouse, or donkey-anti-rat-dylight 549 and donkey-anti-rabbit-dylight 649 IgG (all from Dianova). Before mounting the slides with Mowiol (Hoechst), nuclei were stained with Yoporo (1:2,000; Invitrogen). The sections were analysed using a Leica TCS-Sp/Leica DM RB confocal laser scanning microscope (Leica Microsystems). Images were processed with the Leica Confocal Software LCS (Version 2.61).

Formalin-fixed pancreata were embedded in paraffin. Sections (3–5-µm thick) were cut and stained with haematoxylin and eosin. Immunohistochemistry was performed on an automated immunostainer (Ventana Medical System) according to the manufacturer's protocol, with minor modifications⁴¹. The antibody panel used included activated caspase 3 (Cell Signaling Technology), Ki67 (DCS Innovative Diagnostik-Systeme), pHP1γ (Abcam) and H3K9me3 (Cell Signaling Technology). Formalin-fixed and paraffin-embedded melanoma were stained using monoclonal mouse-anti human Ki67 (clone MIB-1; DAKO) or the CINtec p16INK4a Cytology Kit (Roche mtm laboratories AG).

SA-β-gal activity assay. Cancer cells were fixed for 15 min at room temperature, and then stained for 16 h at 37 °C using the β-Galactosidase Staining Kit (United States Biological). SA-β-gal-positive and -negative cells were then counted using a Zeiss Axiovert 200 microscope (Zeiss). In some experiments, the cells were counterstained for synaptophysin by immunofluorescence, and synaptophysin-SA-β-gal double-positive cells were counted.

Cell cycle analysis. After treatment of β-cancer cells, cell cycle analysis was performed using the BD Pharmingen FITC-BrdU Flow Kit according to the manufacturer's protocol (BD Biosciences). The samples were analysed by flow cytometry on a LSR II from Becton Dickinson, and the following cell cycle phases were determined as a percentage of the total population: sub-G1 (apoptotic cells), G1/G0 (2n, BrdU-negative), S (2n to 4n, BrdU-positive) and G2/M phase (4n, BrdU-negative).

Western blotting. After treatment, cancer cells were lysed in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5% SDS, 1 mM NaF, 1 mM Na₃VO₄ and 0.4% β-mercaptoethanol) containing a protease inhibitor cocktail and a phosphatase inhibitor cocktail (PhosSTOP, Roche Diagnostics). Alternatively, cytoplasmic protein extracts were obtained from the cell cultures using the NE-PER extraction kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Before use, protease inhibitors and PhosSTOP were added to the lysis buffers CER I and NER. After determination of protein content by the bicinchoninic acid assay (Thermo Fisher Scientific), proteins were resolved by 12% SDS-PAGE or by Mini-PROTEAN TGX Precast Gels (4–15%; BioRad), transferred

onto a polyvinylidene difluoride (PVDF) membrane and blocked with 3% non-fat milk in TBS/0.1% Tween 20 (TBST) as described⁴². The membrane was incubated with anti-p16INK4a (1:1,000; Santa Cruz), anti-Rb (Ab-780) (1:1,000), anti-Rb(phospho-Ser-795) (1:1,000; both from SAB Signalway Antibody), anti-CXCL9 (1:2,000), anti-CXCL10 (1:2,000; both from R&D Systems), or anti- β -actin antibody (1:1,000; BioVision). After washing with TBST and subsequent blocking, the blots were incubated with goat anti-mouse horseradish peroxidase (HRP)-conjugated antibody or with goat anti-rabbit HRP-conjugated antibody (1:3,000; Cell Signaling Technology), washed again, and antibody binding was detected with the ECL detection reagent (Amersham). Immunoreactive bands were quantified using the ImageJ software (National Institutes of Health), and the phospho-Rb/Rb ratio of the samples was calculated.

Chromium release assay. 2.5×10^6 target cells were labelled with 250 μ Ci (9.25 MBq) 51 NaCr (Hartmann Analytic) at 37 °C for 1.5 h. After incubation, cells were washed and plated into microtitre round bottom plates at 1×10^4 cells per well. Effector cells were added at different effector to target ratios and incubated at 37 °C for 4 h. Spontaneous release in the absence of effector cells was always less than 30% of the maximal release induced by Triton X-100 (1%). After incubation, 50 μ l supernatant per well was mixed with 200 μ l scintillation cocktail (Ultima Gold, PerkinElmer) and measured in a liquid scintillation counter (MicroBeta, PerkinElmer).

Gene-expression analysis. The gene expression of β -cancer cells and different mouse and human cancer cell lines was analysed after RNA purification and reverse transcription by PCR essentially as described⁴³. The following primers were used: mouse *Cdkn2a* (*p16INK4a*) (146 base pairs (bp)): sense TTGCCC ATCATCATCACCT and antisense GGGTTTTCTTGGTGAAGTTCG; mouse *Ifng1* (474 bp): sense AGGAGGAAGAAGGAAGAACAG and antisense TACCACAGAGAGCAAGGAC; mouse *Ifng2* (533 bp): sense TCATACACTTC TCCCTCCC and antisense CACATCATCTCGCTCCTTTTC; mouse *Tnfrsf1a* (531 bp): sense TTCCCCTCCTACCTTCTCTCT C and antisense TGCCCTTTT CACTCCCTG; mouse *Hprt* (90 bp): sense TCCTCCTCAGACCGCTTTT and antisense CCTGGTTCATCATCGCTAATC; mouse *Eef1a1* (119 bp): sense ACACGTAGATTCCGGCAAGT and antisense AGGAGCCCTTCCCCATCTC; mouse *Cxcl9* (80 bp): sense TTTTCTCTTGGGCATCATCTT and antisense AGCATCGTGATTCCTTATCACT; mouse *Cxcl10* (111 bp): sense GCTGCCG TCATTTTCTGC and antisense TCTCACTGGCCGTCATC; human *IFNGR1* (417 bp): sense TCTCCTCTTCTCCTACCCC and antisense ATTTGCTTCTCC

TCCTTTCTG; human *IFNGR2* (558 bp): sense GTACACAGATCACAGC AACAG and antisense TCAGGACCAGGAAGAAACAG; human *TNFRSF1A* (580 bp): sense GGACAGGGAGAAGAGAGATAG and antisense GAGGAGGG ATAAAAGGCCAAAG.

Statistics. If not otherwise stated, data were expressed as arithmetic means \pm s.e.m., and statistical analyses were made by unpaired *t*-test, or analysis of variance using Dunnett's or Tukey's test as a post hoc test, where appropriate. $P < 0.05$ was considered statistically significant.

- Hanahan, D. Heritable formation of pancreatic β -cell tumours in transgenic mice expressing recombinant insulin/simian virus 40 oncogenes. *Nature* **315**, 115–122 (1985).
- Pfeffer, K. *et al.* Mice deficient for the 55 kd tumor necrosis factor receptor are resistant to endotoxic shock, yet succumb to L. monocytogenes infection. *Cell* **73**, 457–467 (1993).
- Förster, I., Hirose, R., Arbeit, J. M., Clausen, B. E. & Hanahan, D. Limited capacity for tolerization of CD4⁺ T cells specific for a pancreatic β cell neo-antigen. *Immunity* **2**, 573–585 (1995).
- Maglione, J. E. *et al.* Transgenic polyoma middle-T mice model premalignant mammary disease. *Cancer Res.* **61**, 8298–8305 (2001).
- Shultz, L. D., Ishikawa, F. & Greiner, D. L. Humanized mice in translational biomedical research. *Nature Rev. Immunol.* **7**, 118–130 (2007).
- Ziegler, A. *et al.* EpCAM, a human tumor-associated antigen promotes Th2 development and tumour immune evasion. *Blood* **113**, 3494–3502 (2009).
- Monks, A. *et al.* Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J. Natl. Cancer Inst.* **83**, 757–766 (1991).
- Keyes, W. M. *et al.* p63 deficiency activates a program of cellular senescence and leads to accelerated aging. *Genes Dev.* **19**, 1986–1999 (2005).
- Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature Genet.* **37**, 1289–1295 (2005).
- Kneilling, M. *et al.* Direct crosstalk between mast cell-TNF and TNFR1-expressing endothelia mediates local tissue inflammation. *Blood* **114**, 1696–1706 (2009).
- Kunder, S. *et al.* A comprehensive antibody panel for immunohistochemical analysis of formalin-fixed, paraffin-embedded hematopoietic neoplasms of mice: analysis of mouse specific and human antibodies cross-reactive with murine tissue. *Toxicol. Pathol.* **35**, 366–375 (2007).
- Hennige, A. M. *et al.* Overexpression of kinase-negative protein kinase C δ in pancreatic β -cells protects mice from diet-induced glucose intolerance and β -cell dysfunction. *Diabetes* **59**, 119–127 (2010).
- Biedermann, T. *et al.* Mast cells control neutrophil recruitment during T cell-mediated delayed-type hypersensitivity reactions through tumor necrosis factor and macrophage inflammatory protein 2. *J. Exp. Med.* **192**, 1441–1452 (2000).

APOBEC3B is an enzymatic source of mutation in breast cancer

Michael B. Burns^{1,2,3,4,*}, Lela Lackey^{1,2,3,4,*}, Michael A. Carpenter^{1,2,3,4}, Anurag Rathore^{1,2,3,4}, Allison M. Land^{1,2,3,4},
 Brandon Leonard^{2,3,4,5}, Eric W. Refsland^{1,2,3,4}, Delshanee Kotandeniya^{2,6}, Natalia Tretyakova^{2,6}, Jason B. Nikas², Douglas Yee²,
 Nuri A. Temiz⁷, Duncan E. Donohue⁷, Rebecca M. McDougale^{1,2,3,4}, William L. Brown^{1,2,3,4}, Emily K. Law^{1,2,3,4}
 & Reuben S. Harris^{1,2,3,4,5}

Several mutations are required for cancer development, and genome sequencing has revealed that many cancers, including breast cancer, have somatic mutation spectra dominated by C-to-T transitions¹⁻⁹. Most of these mutations occur at hydrolytically disfavoured¹⁰ non-methylated cytosines throughout the genome, and are sometimes clustered⁸. Here we show that the DNA cytosine deaminase APOBEC3B is a probable source of these mutations. *APOBEC3B* messenger RNA is upregulated in most primary breast tumours and breast cancer cell lines. Tumours that express high levels of *APOBEC3B* have twice as many mutations as those that express low levels and are more likely to have mutations in *TP53*. Endogenous APOBEC3B protein is predominantly nuclear and the only detectable source of DNA C-to-U editing activity in breast cancer cell-line extracts. Knockdown experiments show that endogenous APOBEC3B correlates with increased levels of genomic uracil, increased mutation frequencies, and C-to-T transitions. Furthermore, induced APOBEC3B overexpression causes cell cycle deviations, cell death, DNA fragmentation, γ -H2AX accumulation and C-to-T mutations. Our data suggest a model in which APOBEC3B-catalysed deamination provides a chronic source of DNA damage in breast cancers that could select *TP53* inactivation and explain how some tumours evolve rapidly and manifest heterogeneity.

Most humans encode a total of 11 polynucleotide cytosine deaminase family members that could contribute to mutation in cancer—APOBEC1, activation-induced deaminase (AID), APOBEC2, APOBEC3 proteins (known as A3A, A3B, A3C, A3D, A3F, A3G and A3H), and APOBEC4. APOBEC2 and APOBEC4 have not shown activity. APOBEC1 and AID are expressed tissue specifically and implicated in cancers of those tissues, hepatocytes and B cells, respectively^{11,12}. We therefore proposed that one or more of the seven APOBEC3 proteins may be responsible for the C-to-T mutations in other human cancers. This possibility is consistent with hybridization¹³ and expression studies¹⁴ (Supplementary Fig. 1).

To identify the contributing APOBEC3 protein, we quantified mRNA levels for each of the 11 family members in breast cancer cell lines (Supplementary Fig. 2). Surprisingly, only *APOBEC3B* mRNA trended towards upregulation. This analysis was expanded to include a total of 38 independent breast cancer cell lines. *APOBEC3B* was upregulated by ≥ 3 s.d. relative to controls in 28 out of 38 lines, with levels exceeding tenfold in 12 out of 38 lines (Fig. 1a and Supplementary Table 1). Of the representative cell lines used, MDA-MB-453, MDA-MB-468 and HCC1569 showed 20-, 21- and 61-fold upregulation, respectively. These results correlate with cell-line

microarray data (Supplementary Fig. 3, Supplementary Tables 2–9 and Supplementary Discussion). *APOBEC3B* upregulation is probably due to an upstream signal transduction event because it is not a frequent site of rearrangement or copy number variation (<http://dbCRID.biolead.org>), and sequencing failed to reveal promoter-activating mutations or CpG islands indicative of epigenetic regulation (Supplementary Fig. 4).

Epitope-tagged APOBEC3B (A3B) localizes to the nucleus of several transfected cell types¹⁵. To determine whether this is also a property of breast cancer lines, a construct encoding A3B fused to enhanced green fluorescent protein (A3B-eGFP) was transfected into MDA-MB-453,

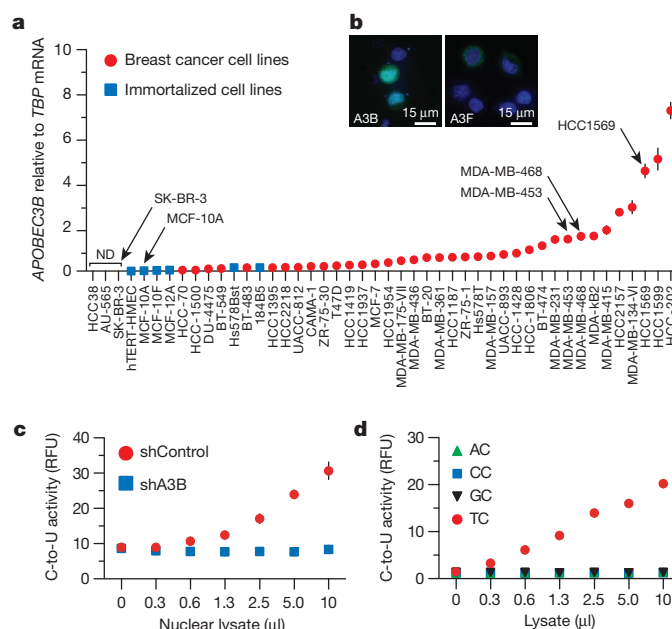


Figure 1 | *APOBEC3B* upregulation and activity in breast cancer cell lines.

a, *APOBEC3B* levels in indicated cell lines. Each point represents the mean of three reactions presented relative to *TBP* (s.d. shown unless smaller than symbol). ND, not detected. **b**, A3B-eGFP or A3F-eGFP localization in MDA-MB-453 cells (nuclei are blue). **c**, Nuclear DNA C-to-U activity in extracts from MDA-MB-453 transduced with shControl or shA3B lentiviruses ($n = 3$; s.d. shown unless smaller than symbol). RFU, relative fluorescence units. **d**, Intrinsic dinucleotide DNA deamination preference of endogenous A3B in extracts from MDA-MB-453 cells ($n = 3$; s.d. smaller than symbols).

¹Biochemistry, Molecular Biology and Biophysics Department, University of Minnesota, Minneapolis, Minnesota 55455, USA. ²Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. ³Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁴Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA.

⁵Microbiology, Cancer Biology and Immunology Graduate Program, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁶Department of Medicinal Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁷In Silico Research Centers of Excellence, Advanced Biomedical Computing Center, Information Systems Program, SAIC-Frederick Inc., Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, USA.

*These authors contributed equally to this work.

MDA-MB-468 and HCC1569 cells. Live cell images showed nuclear localization of A3B-eGFP, in contrast to the cytoplasmic localization of an A3F-eGFP construct (Fig. 1b and Supplementary Fig. 5). Corroborating data were obtained for haemagglutinin (HA)-tagged proteins (Supplementary Fig. 5). To study endogenous A3B subcellular compartmentalization and activity, we used a fluorescence-based DNA C-to-U assay. We first found that nuclear, but not cytoplasmic, fractions of several breast cancer cell lines contain a robust DNA editing activity, which could be ablated by *APOBEC3B* knockdown (Fig. 1c and Supplementary Figs 6 and 7). Similar results were obtained with an independent knockdown construct (not shown). Protein extracts were then used to assess the local dinucleotide deamination preference of endogenous A3B. Similar to retroviral hypermutation signatures caused by A3B overexpression¹⁶, endogenous A3B showed a strong preference for editing cytosines in the TC dinucleotide context (Fig. 1d and Supplementary Fig. 6). No deaminase activity was observed for extracts from MCF-10A (A3B^{low}) or SK-BR-3 (A3B^{null}) cells, although it could be conferred by transient A3B transfection (Supplementary Fig. 8). Both A3B-HA and A3A-HA could elicit measurable TC-to-TU activity in lysates from transfected HEK293T cells (Supplementary Fig. 9). However, because *APOBEC3A* mRNA is myeloid lineage-specific¹⁷ and non-detectable in breast cancer cell lines (Supplementary Figs 1 and 2), our expression and activity studies indicated that A3B may be the only enzyme poised to deaminate breast cancer genomic DNA.

To address whether endogenous A3B damages genomic DNA, we used a combination of biophysical and genetic assays. We first used a mass spectrometry-based approach to quantify levels of genomic uracil in MDA-MB-453 and HCC1569 cells with high levels of endogenous A3B versus knockdown levels of A3B (short hairpin RNA (shRNA) control versus shRNA against *APOBEC3B* (shA3B)) (Fig. 2a and Supplementary Fig. 10). Genomic uracil loads decreased by 30% in HCC1569 cells expressing shA3B and by 70% in MDA-MB-453 cells, in which knockdown was stronger (Fig. 2b and Supplementary Fig. 10). Although these relative differences may seem modest—10 and 20 uracils per megabases (Mb), respectively—this equates to 30,000 and 60,000 A3B-dependent uracils per haploid genome. The actual number of pro-mutagenic uracils may be even higher because several repair pathways may concurrently function to limit this damage.

Second, we used a thymidine kinase-positive (TK^{plus}) to -negative (TK^{minus}) fluctuation analysis¹⁷ to determine whether upregulated

A3B and increased uracil loads lead to higher levels of mutation. MDA-MB-453 and HCC1569 cells were engineered to express the herpes simplex virus type 1 TK gene, which confers sensitivity to the drug ganciclovir. TK^{plus} lines were transduced with shA3B or shControl constructs, and limiting dilution was used to generate single-cell subclones. Expanded subclones were subjected to ganciclovir selection and resistant cells were grown to visible colonies, which showed that cells with upregulated A3B accumulate 3–5-fold more mutations (Fig. 2c and Supplementary Fig. 10).

Third, differential DNA denaturation PCR (3D-PCR)^{17,18} was used to determine whether C-to-T transition mutations accumulate differentially at three genomic loci in A3B^{low} and A3B^{high} pools of HCC1569 cells. This technique enables qualitative estimates of genomic mutation within a population of cells because DNA sequences with higher A/T content amplify at lower denaturation temperatures than parental sequences. Lower temperature amplicons were observed for *TP53* and *c-MYC*, but not *CDKN2B* (Fig. 2d and Supplementary Fig. 10). These amplicons were cloned and sequenced, and more C-to-T transition mutations were observed in A3B^{high} compared with A3B^{low} samples (Fig. 2d and Supplementary Fig. 10). *TP53* and *c-MYC* appeared more mutable than *CDKN2B*, suggesting that all genomic regions are not equally susceptible to enzymatic deamination. Other base substitution mutations were rare, and some C-to-T transitions were still evident in the A3B^{low} samples, possibly owing to residual deaminase activity and/or amplification of spontaneous events.

To address whether A3B triggers other cancer hallmarks¹⁹, we tried and failed to stably express A3B in several epithelial cell lines. We therefore constructed a panel of HEK293 clones with doxycycline (Dox)-inducible A3B, A3B(E68A/E255Q), A3A, or A3A(E72A) eGFP fusions. As measured by flow cytometry, A3-eGFP levels were barely detectable without Dox and induced in nearly 100% of cells with Dox (Supplementary Fig. 11). A3A overexpression caused rapid S-phase arrest, cytotoxicity and γ -H2AX focus formation, as reported previously²⁰ (Fig. 3a–c and Supplementary Fig. 11). In comparison, A3B induction caused a delayed cell cycle arrest, a more pronounced formation of abnormal anucleate and multinucleate cells, and eventual cell death (Fig. 3a, b and Supplementary Fig. 11). A3B induction also caused γ -H2AX focus formation, DNA fragmentation, as evidenced by visible comets, and C-to-T mutations (Fig. 3c–e). A3B catalytic activity, as evidenced by the glutamate mutants, was required for the induction of these cancer phenotypes.

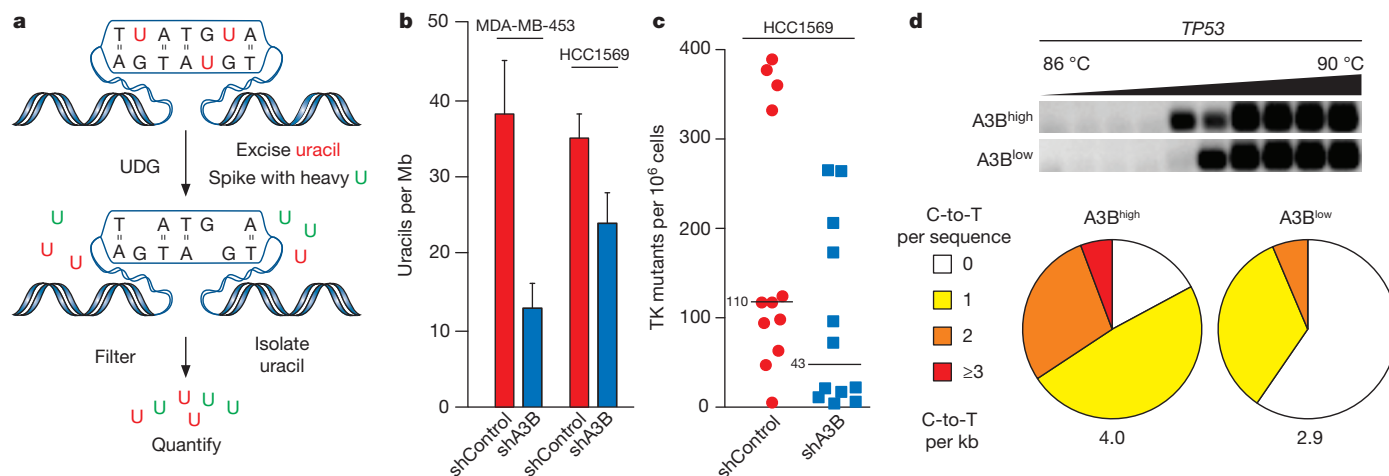


Figure 2 | A3B-dependent uracil lesions and mutations in breast cancer genomic DNA. **a**, Workflow for genomic uracil quantification by high-performance liquid chromatography–tandem mass spectrometry (HPLC-ESI-MS/MS). **b**, Average uracil loads in the indicated cell lines ($n = 3$; errors, s.d.). **c**, Dot plot representing thymidine kinase mutant frequencies of HCC1569

subclones expressing shControl or shA3B. Each dot corresponds to one subclone. Medians are labelled. **d**, Agarose gel and mutation analysis of *TP53* 3D-PCR amplicons from HCC1569 cells expressing shControl (A3B^{high}) or shA3B (A3B^{low}) ($n \geq 35$ sequences per condition). See Supplementary Fig. 10 for further data.

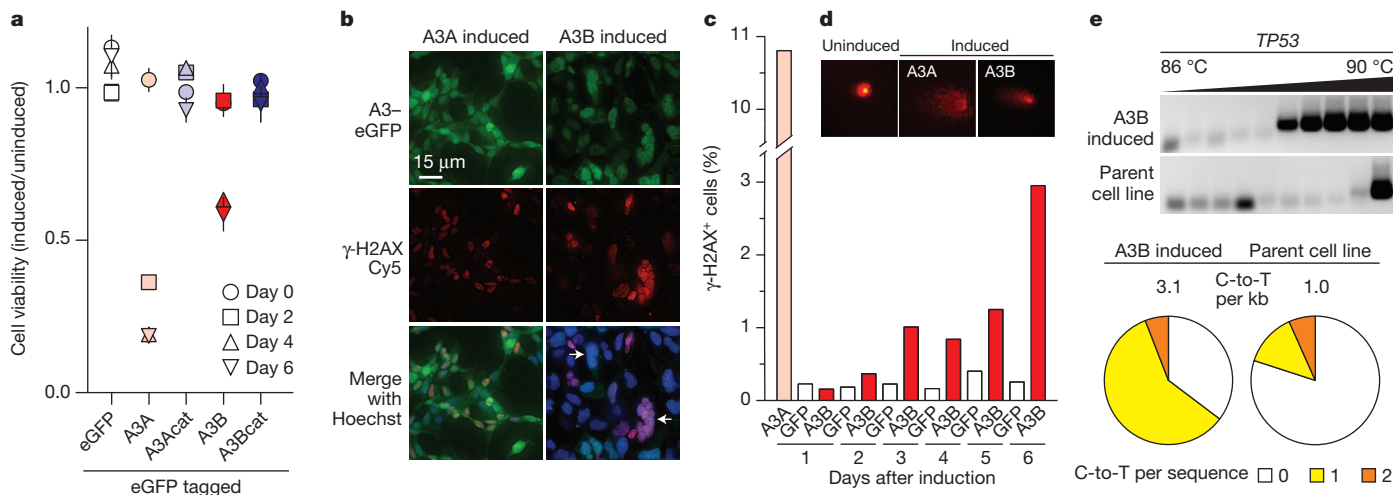


Figure 3 | Cancer phenotypes triggered by inducing A3B overexpression. **a**, Cell viability at indicated times after induction (mean and s.d. for $n = 3$ per condition). A3Acat and A3Bcat denote catalytically defective glutamate mutants. **b**, **c**, Representative fields of cells imaged for γ -H2AX and A3A-eGFP (1 day) or A3B-eGFP (3 days) after induction, and γ -H2AX quantification.

We next asked whether our cell-based results could be extended to primary tumours. First, we quantified mRNA levels for each of the 11 family members in 21 randomly chosen breast tumour specimens, in parallel with matched normal tissue procured simultaneously from an adjacent area or the contralateral breast. Only *APOBEC3B* was expressed preferentially in tumours ($P = 0.0003$) (Supplementary Fig. 12). We confirmed this analysis by measuring *APOBEC3B* levels in 31 additional tumour/normal matched tissue sets. In total, *APOBEC3B* was upregulated by ≥ 3 s.d. in 20 out of 52 tumours in comparison to the patient-matched normal tissue mean, and in 44 out of 52 tumours in comparison to the reduction mamplasty tissue mean (Fig. 4a; $P = 7.1 \times 10^{-7}$ and $P = 2 \times 10^{-5}$, respectively; patient information in Supplementary Table 10). These are underestimates because tumour specimens have varying fractions of non-*APOBEC3B*-expressing normal cells. Some of the matched 'normal' samples may also be contaminated by tumour cells, as judged by the mean levels in mamplasty samples (Fig. 4a; $P = 0.002$). The related deaminase, *APOBEC3G*, was not expressed differentially in these samples, indicating that these observations are not due to immune cells known to express several *APOBEC3* proteins¹⁴ ($P = 0.591$). Similar results were obtained by quantifying RNA-sequencing data for independent matched tumour and normal pairs²¹, with $\sim 50\%$ showing upregulated A3B (defined as tumours with A3B levels > 3 s.d. above the mean of the normal matched samples; $P < 0.0001$) (Fig. 4b).

Finally, we assessed the effect of A3B on the breast tumour genome by correlating the deamination signature of A3B *in vitro* and the somatic mutation spectra accumulated during tumour development *in vivo*. Using a series of single-stranded DNA substrates varying only at the immediate 5' or 3' position relative to the target cytosine (underlined), we found that recombinant A3B prefers $\text{TC} > \text{CC} > \text{GC} = \text{AC}$ and $\text{CA} > \text{CT} = \text{CC}$ (Supplementary Fig. 13). These local sequence preferences were then compared to the expected distribution of cytosine in the human genome and the reported C-to-T mutation profiles for melanoma²², liver²³ and breast^{8,9,21} tumours. Consistent with a spontaneous origin, the C-to-T frequency is low in liver tumours ($\sim 20\%$) and mutational events appear random (Fig. 4c, d). As expected, C-to-T frequencies are high in melanomas ($\sim 80\%$) and focused at dipyrimidines consistent with ultraviolet-induced lesions and subsequent error-prone lesion bypass synthesis (Fig. 4c, d). Interestingly, the C-to-T frequency was intermediate in three

Abnormal, multinuclear clusters are typical of induced A3B-eGFP (white arrows). **d**, Representative images of A3-induced DNA comets (original magnification, $\times 400$). **e**, C-to-T mutations in *TP53* detected by sequencing 3D-PCR products 4 days after induction ($n > 12$ sequences per condition).

independent breast tumour data sets ($\sim 40\%$) and largely focused at trinucleotides that mimic the preferred sites for A3B-dependent DNA deamination *in vitro* (Fig. 4c, d and Supplementary Fig. 13). The availability of both high-throughput RNA sequencing (RNA-seq) and somatic mutation data²¹ also enabled the establishment of strong positive correlations between *APOBEC3B* expression levels and the C-to-T mutation load, overall base substitution mutation load, and *TP53* inactivation (Fig. 4e–g). Importantly, tumours expressing high A3B levels have twice as many mutations (Fig. 4e, f and Supplementary Fig. 14). This equates to 10 C-to-T and 30 total mutations per exome, or approximately 1,000 and 3,000 mutations per genome, being attributable to A3B.

Taken together, we conclude that A3B is an important mutational source in breast cancer accounting for C-to-T mutation biases and increased mutational loads. Moreover, the disproportional increase in overall base substitutions indicates that some of these other patterns may be due to further processing of U/G mispairs by 'repair' enzymes into transitions, transversions and DNA breaks that could precipitate chromosomal rearrangements (model in Supplementary Fig. 15 with similarities to AID-dependent antibody diversification mechanisms²⁴). Future work is needed to understand A3B regulation and the potential interaction with other oncogenes and tumour suppressors. For example, although several common breast cancer markers do not correlate with *APOBEC3B* upregulation, a mechanistic linkage between increased *APOBEC3B* and inactivated *TP53* is evident in primary tumour data and cell lines (Fig. 4g and Supplementary Fig. 16). *TP53* inactivation may be required to allow cells to bypass DNA damage checkpoints triggered by A3B.

This is the first study, to our knowledge, to demonstrate upregulation of the DNA deaminase A3B in breast cancer and reveal it as a considerable source of enzymatic mutation. Conceptually supportive of the original mutator hypothesis²⁵, A3B-catalysed genomic DNA deamination could provide genetic fuel for cancer development, metastasis, and even therapy resistance. We propose that A3B is a dominant underlying factor that contributes to tumour heterogeneity by broadly affecting several pathways and phenotypes. A3B may represent a new marker for breast cancer and a strong candidate for targeted intervention, especially given its non-essential nature²⁶. A3B inhibition may decrease the rate of tumour evolution and stabilize the targets of existing therapeutics.

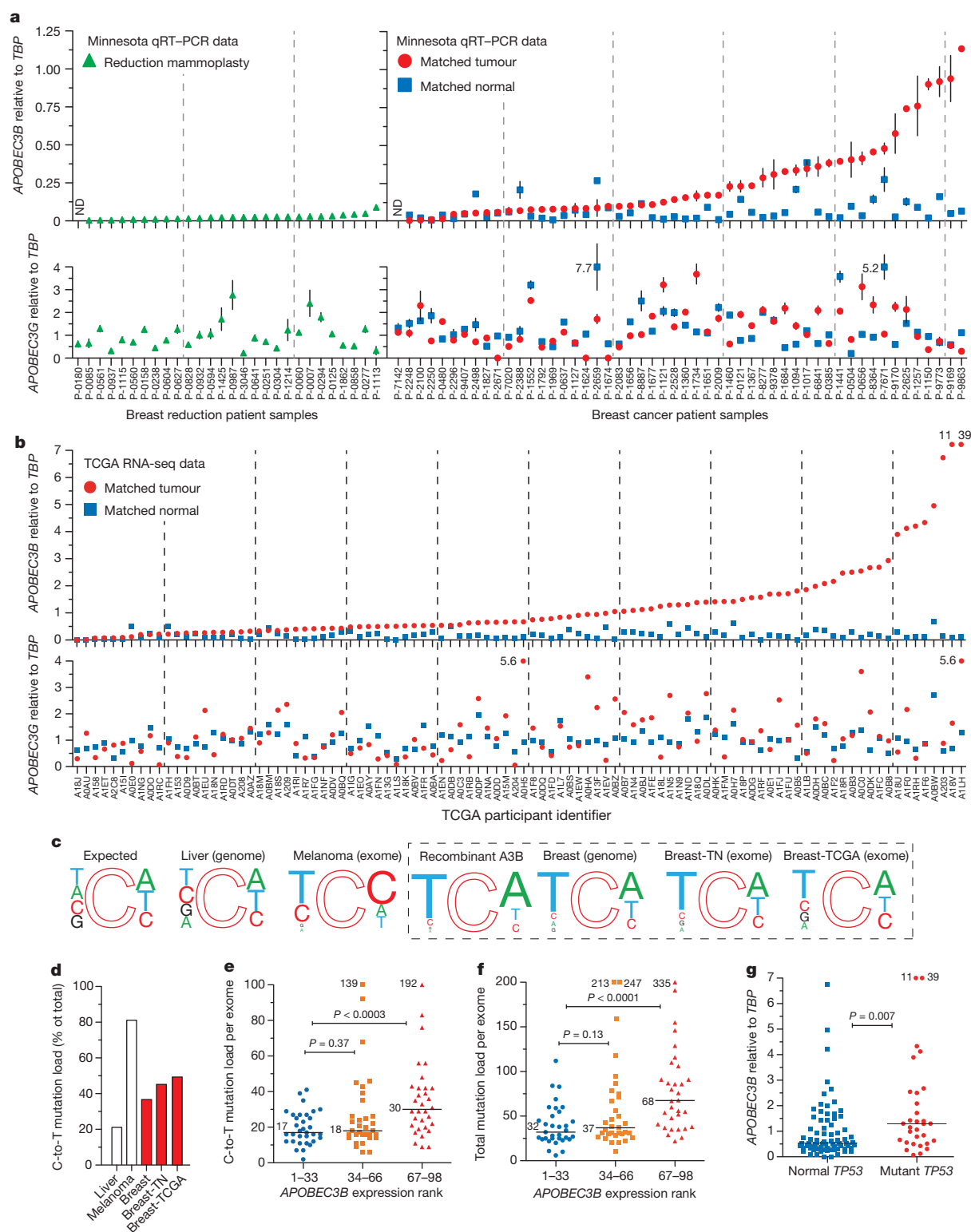


Figure 4 | *APOBEC3B* upregulation and mutation in breast tumours. **a**, *APOBEC3B* and *APOBEC3G* mRNA levels in the indicated tissues. Each symbol represents the mean mRNA level of three quantitative PCR with reverse transcription (qRT-PCR) reactions, presented relative to *TBP* (s.d. shown unless smaller than symbol). **b**, RNA-seq data for *APOBEC3B* and *APOBEC3G* in the indicated samples. TCGA, The Cancer Genome Atlas. **c**, Local sequence contexts for all genomic cytosines (expected), cytosines deaminated by recombinant A3B (Supplementary Fig. 13), and observed

C-to-T transitions in the indicated cancers. Font size is proportional to nucleotide frequency. TN, triple-negative. **d**, Percentage of C-to-T mutations in the indicated tumours. **e**, **f**, C-to-T (**e**) and total (**f**) mutation counts for tumours in **b** grouped into lower, middle and upper thirds based on *APOBEC3B* levels (medians are labelled). **g**, Relationship between *APOBEC3B* level (RNA-seq data) and *TP53* status for tumours in **b**. Off-scale values in **a**, **b**, **e**–**g** are indicated numerically; *P* values in **e**–**g** are from Mann–Whitney *U* test.

METHODS SUMMARY

Flash-frozen tissues were obtained from the University of Minnesota Tissue Procurement Facility. Availability of both tumour and matched normal tissue was the only selection criteria. Mammary reduction samples were used as non-cancer controls. These studies were performed in accordance with Institutional Review Board (IRB) guidelines (IRB study number 1003E78700). The breast cancer cell line panel 30-4500K was obtained from the ATCC and cultured as recommended. RNA isolation, complementary DNA synthesis, and quantitative PCR procedures were performed as reported¹⁴ (Supplementary Table 11). Knockdown and control shRNA constructs were obtained from Open Biosystems. Microscopy, cellular fractionation and deaminase activity assays were done as described^{15,17}. Genomic uracil was quantified by treating DNA samples with uracil DNA glycosylase, purifying the nucleobase away from the remaining DNA, and analysing the samples by mass spectrometry. The TK and 3D-PCR mutation assays have been described and were modified for use with breast cancer cell lines¹⁷. Dox-inducible cells were obtained from Invitrogen and stable derivatives were created with the indicated constructs. These lines were analysed for cell cycle arrest using propidium iodide staining and cell viability with crystal violet staining and the MTS assay. DNA damage was measured by the comet assay and by flow cytometry and microscopy of cells immunostained for γ -H2AX. Recombinant A3B195-382-mycHis was purified and used for deamination kinetics as described²⁷ using 5'-ATTATTATTATNCNAATGGATTTATTTATTTATTTATTTATTT-6-FAM (NCA and TCN for 5' and 3' preference experiments, respectively). The somatic single-nucleotide mutation frequencies with local sequence contexts were determined by compiling published primary tumour genomic, exomic or RNA sequencing data^{8,9,21–23}. Potential mechanistic overlap with hydrolytic deamination of 5-methyl-cytosines was avoided by excluding CpG dinucleotides from mutational preference calculations.

Full Methods and any associated references are available in the online version of the paper.

Received 9 February; accepted 24 December 2012.

Published online 6 February 2013.

- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Jones, S. *et al.* Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).
- Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
- Kumar, A. *et al.* Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc. Natl Acad. Sci. USA* **108**, 17087–17092 (2011).
- Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331**, 435–439 (2011).
- Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
- Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).
- Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
- Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
- Ehrlich, M., Norris, K. F., Wang, R. Y., Kuo, K. C. & Gehrke, C. W. DNA cytosine methylation and heat-induced deamination. *Biosci. Rep.* **6**, 387–393 (1986).
- Pavri, R. & Nussenzweig, M. C. AID targeting in antibody diversity. *Adv. Immunol.* **110**, 1–26 (2011).
- Yamanaka, S. *et al.* Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc. Natl Acad. Sci. USA* **92**, 8483–8487 (1995).
- Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002).
- Refsland, E. W. *et al.* Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. *Nucleic Acids Res.* **38**, 4274–4284 (2010).
- Lackey, L. *et al.* APOBEC3B and AID have similar nuclear import mechanisms. *J. Mol. Biol.* **419**, 301–314 (2012).
- Albin, J. S. & Harris, R. S. Interactions of host APOBEC3 restriction factors with HIV-1 *in vivo*: implications for therapeutics. *Expert Rev. Mol. Med.* **12**, e4 (2010).
- Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J. & Harris, R. S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nature Struct. Mol. Biol.* **17**, 222–229 (2010).
- Suspène, R. *et al.* Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc. Natl Acad. Sci. USA* **108**, 4858–4863 (2011).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* **12**, 444–450 (2011).
- The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- Wei, X. *et al.* Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature Genet.* **43**, 442–446 (2011).
- Zhang, J. *et al.* International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database* **2011**, bar026 (2011).
- Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
- Loeb, L. A., Springgate, C. F. & Battula, N. Errors in DNA replication as a basis of malignant changes. *Cancer Res.* **34**, 2311–2321 (1974).
- Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R. & Eichler, E. E. Population stratification of a common APOBEC gene deletion polymorphism. *PLoS Genet.* **3**, e63 (2007).
- Carpenter, M. A. *et al.* Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem.* **287**, 34801–34808 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Hultquist and R. Vogel for statistics, T. Hwang for bioinformatic assistance, V. Polunovsky for hTERT-HMECs, V. Simon for shRNA, S. Kaufmann, C. Lange and D. Largaespada for consultation, and the Masonic Cancer Center Breast Cancer Research Fund for purchasing the ATCC breast cancer panel. Tissues were obtained from the Masonic Cancer Center Tissue Procurement Facility, which is part of BioNet, supported by the Academic Health Center and National Institutes of Health (NIH) grants P30 CA77598 (D.Y.), P50 CA101955 (D. Buchsbaum) and KL2 RR033182 (B. Blazar). M.B.B. was supported in part by a Cancer Biology Training Grant (NIH NCI T32 CA009138) and a Department of Defense Breast Cancer Research Program Predoctoral Fellowship (BC101124). L.L. was supported in part by a National Science Foundation Predoctoral Fellowship and by a position on the Institute for Molecular Virology Training Grant NIH T32 AI083196. M.A.C. was supported by an NIH postdoctoral fellowship (F32 GM095219). A.M.L. was supported by a CIHR postdoctoral fellowship. E.W.R. was supported by a position on the Institute for Molecular Virology Training Grant NIH T32 AI083196 and subsequently by an NIH predoctoral fellowship (F31 DA033186). Computational analyses (N.A.T. and D.E.D.) were supported by federal funds from the National Cancer Institute, NIH, CBIIT/caBIG ISRC yellow task 09-260. The Harris laboratory was supported in part by NIH R01 AI064046, NIH P01 GM091743, the Children's Cancer Research Fund, and a seed grant from the University of Minnesota Clinical and Translational Science Institute (supported by NIH 1UL1RR033183).

Author Contributions R.S.H. conceived and managed the overall project. M.B.B. assisted R.S.H. with experimental design, project management and manuscript preparation. M.B.B., E.W.R. and B.L. generated mRNA expression profiles; L.L. and E.K.L. performed microscopy; L.L. and A.R. performed biochemical fractionations and DNA deaminase assays; M.B.B. performed uracil quantifications; A.M.L. performed thymidine kinase fluctuations; A.R. generated 3D-PCR sequences; and L.L., A.M.L., A.R. and M.A.C. determined the effect of induced A3B overexpression. M.A.C. performed deaminase assays with recombinant protein; and M.A.C. and D.K. assisted with the HPLC-ESI-MS/MS set up. N.T. was involved in HPLC-ESI-MS/MS method development. J.B.N. conducted the search and performed the bioinformatic analysis of the microarray data and developed the normalization algorithm for this analysis. N.A.T., D.E.D. and M.B.B. contributed bioinformatic analyses. All authors contributed to manuscript revisions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S.H. (rs@uminn.edu).

METHODS

RNA isolation, cDNA synthesis and qRT-PCR. Matched tumour/normal breast tumours and mammary reduction samples from the University of Minnesota Tissue Procurement Facility and breast cancer cell lines 30-4500K from the ATCC were used for RNA isolation, cDNA synthesis and qPCR as described¹⁴. Tissue RNA was from 100 mg flash-frozen tissue disrupted by a 2-h water bath sonication in 1 ml of Qiazol Lysis Reagent (RNeasy, Qiagen). Cell RNA was made using QiaShredder (RNeasy, Qiagen). qPCR was performed on a Roche Lightcycler 480 instrument. The housekeeping gene *TBP* was used for normalization. Statistical analyses for matched tissues were done using the Wilcoxon signed-rank test, and unmatched sets with the Mann-Whitney U-test (Graphpad Prism). Primer and probe sequences are listed in Supplementary Table 11.

Knockdown constructs. *APOBEC3B* shRNA and shControl lentiviral constructs were from Open Biosystems (TRCN0000157469, TRCN0000140546 and scramble). Knockdown levels ranged from 80% to 95% by qRT-PCR. Helper plasmids pΔ-NRF, containing HIV-1 *gag*, *pol*, *rev* and *tat* genes, and pMDG, containing the VSV-G *env* gene, were co-transfected in HEK293T cells. Cell-free supernatants were collected and concentrated by centrifugation (14,000g for 2 h). Stable transductants were selected with puromycin (1 μg ml⁻¹).

Cell fractionation and DNA deaminase activity assays. Subcellular activity analysis and dinucleotide preferences were measured as described^{27,28}. In brief, cellular fractionation was performed by syringe treatment of 10⁷ cells in 0.5 ml of hypotonic buffer²⁸. Nuclei were lysed by sonication in lysis buffer (25 mM HEPES, pH 7.4, 250 mM NaCl, 10% glycerol, 0.5% Triton X-100, 1 mM EDTA, 1 mM MgCl₂ and 1 mM ZnCl₂). Anti-histone H3 (1:2,000; Abcam) and anti-tubulin (1:10,000; Covance) followed by anti-mouse 800 or anti-rabbit 680 (1:5,000; Licor) immunoblots were used to assess fractionation. Lysates were tested in a fluorescence-based deaminase activity assay¹⁷. Dilutions were incubated 2 h at 37 °C with a DNA oligonucleotide 5'-(6-FAM)-AAA-TTC-TAA-TAG-ATA-ATG-TGA-(TAMRA). Fluorescence was measured on SynergyMx plate reader (BioTek). Local dinucleotide preferences in extracts were analysed similarly using 5'-AC, CC, GC or TC at the NN position of 5'-(6-FAM)-ATA-ANN-AAA-TAG-ATA-AT-(TAMRA).

Genomic uracil quantifications. Genomic DNA was prepared from shA3B- or shControl-transduced cells cultured for 21 days. Samples were spiked with heavy (+6)-labelled uracil (¹³C and ¹⁵N; Cambridge Isotopes) and treated with uracil-DNA glycosylase (NEB). Uracil was purified using 3,000 molecular mass cut-off columns (Pall Scientific) and solid-phase extraction (Carbograph, Grace). Samples were resuspended in water containing 0.1% formic acid. Analyses were performed on a capillary HPLC-ESI-MS/MS (Thermo-Finnigan Ultra TSQ mass spectrometer, Waters nanoACQUITY HPLC). The mass spectrometer was operated in positive ion mode, with 3.0 kV typical spray voltage, 250 °C capillary temperature, 67 V tube lens offset, and nitrogen sheath gas (25 counts). Argon collision gas was used at 146.7 mPa. Tandem mass spectrometry analyses were performed with a scan width of 0.4 *m/z* and scan time of 0.1 s. The Hypercarb HPLC column (0.5 mm × 100 mm, 5 μm, Thermo Scientific) was maintained at 40 °C and a flow rate of 15 μl min⁻¹. Solvents were 0.1% formic acid and acetonitrile. A linear gradient of 0–8% acetonitrile in 8 min was used, followed by an increase to 80% acetonitrile over 7 min. Uracils eluted at 11.5 min. Selected reaction monitoring was conducted with collision energy of 20 V using the transitions: *m/z* 113.08 [M⁺H]⁺→70.08 [M-CONH]⁺ and *m/z* 96.08 [M-NH₂]⁺ for uracil, whereas the internal standard ([¹⁵N-2, ¹³C-4]-uracil) was monitored by the transitions *m/z* 119.08 [M⁺H]⁺→*m/z* 74.08 [M-CONH]⁺ and *m/z* 101.08 [M-NH₂]⁺, respectively. Internal standards were used for quantification.

TK fluctuations. TK-neo was introduced into MDA-MB-453 and HCC1569 cells as described¹⁷. TK^{plus} cells were transduced with shA3B or shControl lentiviruses and subcloned by limiting dilution. One-million cells from each expanded subclone population were subjected to ganciclovir and incubated until colonies outgrew. Frequencies were determined by applying the method of the median²⁹.

3D-PCR and sequencing. DNA was collected from Ugi-expressing³⁰ T-REx-293 clones or HCC1569 cells transduced with shA3B or shControl lentiviruses. 3D-PCR was done using Taq (Denville Scientific) as described¹⁷. Primers sequences are available on request. PCR products were analysed by gel electrophoresis with ethidium bromide, PCR purified (Epoch), blunt-end cloned into pJET (Fermentas), sequenced with T7 primer (BMGC), and aligned and analysed with Sequencer software (Gene Codes Corporation).

Cell cycle experiments. T-REx-293 cells (Invitrogen) were transfected with pcDNA5/TO A3-GFP using TransIT-LT1 (Mirus) followed by clone selection using hygromycin. Cells were induced with 1 μg ml⁻¹ Dox (MP Biomedicals 198955) for the indicated times then trypsinized and fixed with 4% paraformaldehyde in PBS. Cell pellets were resuspended in 0.1% Triton X-100, 20 μg ml⁻¹ propidium iodide and 40 μg ml⁻¹ RNase A (Qiagen) in PBS for 30 min and the DNA content and GFP induction were measured by flow cytometry (BD Biosciences FACS Canto II) and analysed with FlowJo and GraphPad Prism.

Cell viability assays. Cells were plated into multiple 96-well plates (2,500 cells per well) and measured at the days indicated. The MTS and PMS reagents were used as directed (Promega, Celltiter Aq 96). Absorbance was measured at 490 nm (PerkinElmer 1420 Victor 3V). The results were normalized to untreated cells. For crystal violet staining wells of a 6-well plate were plated with 2 × 10⁵ cells. Half of the wells were induced with 1 μg ml⁻¹ Dox. A crystal violet (0.5%), methanol (49.5%), water (50%) solution was used to stain cells after 7 days.

DNA damage experiments. Flow cytometric analysis of γ-H2AX foci was adapted³¹. Fixed cells were incubated overnight in 0.2% Triton X-100, 1% BSA in PBS (blocking buffer) with 1:100 rabbit anti-γ-H2AX (Bethyl A300-081A). Secondary incubation was with goat anti-rabbit TRITC (Jackson 111025144) for 3 h before flow cytometry (BD Biosciences FACS Canto II) and analysis (FloJo and GraphPad). For microscopy, HEK293 cells were induced with 1 μg ml⁻¹ of Dox before fixation with 4% paraformaldehyde and incubation with 1:50 anti-γ-H2AX conjugated to Alexa 647 (Cell Signaling 20E3) in blocking buffer for 3 h. The cells were stained with 0.1% Hoechst dye and imaged at ×20 or ×60 (Deltavision) and deconvolved (SoftWoRx, Applied Precision).

Comet assays. As described³², microscope slides were coated with 1.5% agarose and dried. Low-melting agarose (0.5% in PBS) was combined 1:1 with HEK293T cells transfected with A3A-eGFP (1 day) or A3B-eGFP (6 day). Ten-thousand cells were added to coated slides and the cells were lysed overnight in 10 mM Tris, 100 mM EDTA, 2.5 M NaCl and 1% Triton X-100. Slides were incubated for 10 min in running buffer (300 mM NaOH, 1 mM EDTA, pH 13.1) then run at 0.75 V cm⁻¹ for 30 min. Gels were neutralized with 0.4 M Tris-HCl, pH 7.5, and treated with RNase A (Qiagen). The microgels were allowed to dry and comets were visualized using propidium iodide.

Bioinformatic analyses. Primary tumour genomic, exomic or RNA sequencing data were obtained from public sources^{8,9,21–23}. Liver tumour genomes had 654,879, melanoma exomes had 2,798, breast tumour genomes had 183,916, breast triple-negative study exomes had 6,964, and TCGA breast tumour exomes had 5,559 total single base substitution mutations. Local contexts were tabulated and presented as weblogo schematics. Complex mutational events and CpG motifs were excluded.

28. Shlyakhtenko, L. S. *et al.* Atomic force microscopy studies provide direct evidence for dimerization of the HIV restriction factor APOBEC3G. *J. Biol. Chem.* **286**, 3387–3395 (2011).
29. Lea, D. E. & Coulson, C. A. The distribution of the numbers of mutants in bacterial populations. *J. Genet.* **49**, 264–285 (1949).
30. Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
31. Huang, X. & Darzynkiewicz, Z. Cytometric assessment of histone H2AX phosphorylation: a reporter of DNA damage. *Methods Mol. Biol.* **314**, 73–80 (2006).
32. Fairbairn, D. W., Olive, P. L. & O'Neill, K. L. The comet assay: a comprehensive review. *Mutat. Res.* **339**, 37–59 (1995).

OTUD7B controls non-canonical NF- κ B activation through deubiquitination of TRAF3

Hongbo Hu¹, George C. Brittain¹, Jae-Hoon Chang¹, Nahum Puebla-Osorio¹, Jin Jin¹, Anna Zal¹, Yichuan Xiao¹, Xuhong Cheng¹, Mikyoung Chang¹, Yang-Xin Fu², Tomasz Zal^{1,3}, Chengming Zhu^{1,3} & Shao-Cong Sun^{1,3}

The non-canonical NF- κ B pathway forms a major arm of NF- κ B signalling that mediates important biological functions, including lymphoid organogenesis, B-lymphocyte function, and cell growth and survival^{1–3}. Activation of the non-canonical NF- κ B pathway involves degradation of an inhibitory protein, TNF receptor-associated factor 3 (TRAF3), but how this signalling event is controlled is still unknown^{1,2}. Here we have identified the deubiquitinase OTUD7B as a pivotal regulator of the non-canonical NF- κ B pathway. OTUD7B deficiency in mice has no appreciable effect on canonical NF- κ B activation but causes hyperactivation of non-canonical NF- κ B. In response to non-canonical NF- κ B stimuli, OTUD7B binds and deubiquitinates TRAF3, thereby inhibiting TRAF3 proteolysis and preventing aberrant non-canonical NF- κ B activation. Consequently, the OTUD7B deficiency results in B-cell hyper-responsiveness to antigens, lymphoid follicular hyperplasia in the intestinal mucosa, and elevated host-defence ability against an intestinal bacterial pathogen, *Citrobacter rodentium*. These findings establish OTUD7B as a crucial regulator of signal-induced non-canonical NF- κ B activation and indicate a mechanism of immune regulation that involves OTUD7B-mediated deubiquitination and stabilization of TRAF3.

Unlike the canonical NF- κ B pathway, which depends on degradation of I κ B α , the non-canonical NF- κ B pathway depends on the inducible processing of p100, a process that not only generates p52 but also leads to nuclear translocation of the p52–RelB dimer². The processing of p100 is triggered through its phosphorylation by the NF- κ B-inducing kinase (NIK) together with a downstream kinase, I κ B kinase- α (IKK- α)^{4,5}. Steady-state non-canonical NF- κ B activity is tightly controlled by TRAF3, which constantly targets NIK to the TRAF2–cIAPs E3 complex for degradation^{6–8}. Activation of non-canonical NF- κ B involves signal-induced TRAF3 proteolysis and NIK accumulation, but how this event is regulated is unknown⁶. The deubiquitinases (DUBs) CYLD and A20 are vital regulators of the canonical NF- κ B pathway, but they are not involved in the regulation of non-canonical NF- κ B signalling^{9,10}. Another DUB, OTUD7B (also called Cezanne), was identified on the basis of its sequence homology with A20 within their ovarian tumour (OTU) domains¹¹. We studied the function of OTUD7B using *Otud7b* knockout (*Otud7b*^{−/−}) mice (Supplementary Figs 1 and 2). In contrast to the postnatal lethality of A20-deficient mice¹², the *Otud7b*^{−/−} mice did not show obvious defects in survival, although they had moderately reduced body weight (Supplementary Fig. 3). The development of B and T cells appeared to be also normal in the *Otud7b*^{−/−} mice (Supplementary Fig. 4). These phenotypes indicate fundamental differences in signalling function between OTUD7B and A20.

The OTUD7B deficiency had no appreciable effect on canonical NF- κ B activation in mouse embryonic fibroblasts (MEFs), bone-marrow-derived macrophages (BMDMs), or B cells (Supplementary Fig. 5). Although OTUD7B has a related molecule, OTUD7A, OTUD7A was detected mainly in the central nervous system (Supplementary Fig. 6), arguing against the possibility of functional redundancies. We next examined the role of OTUD7B in regulating NF- κ B activation

by three well-characterized non-canonical NF- κ B-stimulating receptors: lymphotoxin- β receptor (LT β R), CD40 and BAFF receptor (BAFFR)¹². OTUD7B deficiency greatly enhanced NF- κ B activation by agonistic anti-LT β R and anti-CD40 antibodies or BAFF (Fig. 1a–c), a result that was not due to differential expression of the receptors (Supplementary Fig. 7). The heightened NF- κ B DNA-binding activity in *Otud7b*^{−/−} cells was associated with a marked increase in nuclear levels of p52 and RelB and concomitant loss of cytoplasmic p100 (Fig. 1d–f). However, OTUD7B deficiency did not promote the activation of canonical IKK or MAP kinases (Supplementary Fig. 8), indicating a specific role for OTUD7B in non-canonical NF- κ B regulation. We also found that OTUD7B expression was induced by the non-canonical NF- κ B signals (Supplementary Fig. 9), indicating a negative feedback function.

Compared with wild-type cells, *Otud7b*^{−/−} cells displayed a markedly higher level of TRAF3 degradation and NIK accumulation in response to the non-canonical NF- κ B inducers (Fig. 2a, b and Supplementary Fig. 10). This effect was specific for the non-canonical NF- κ B pathway, as OTUD7B deficiency did not alter the fate or function of TRAF3 in the type I interferon (IFN-I) pathway (Supplementary Figs 11 and 12)^{13,14}. The loss of OTUD7B promoted anti-CD40-stimulated TRAF3 ubiquitination (Fig. 2c). Although the recombinant OTU domain of OTUD7B preferentially hydrolyses K11-linked ubiquitin chains *in vitro*¹⁵, we did not detect appreciable levels of K11 ubiquitination of TRAF3 (Supplementary Fig. 13), which might be due

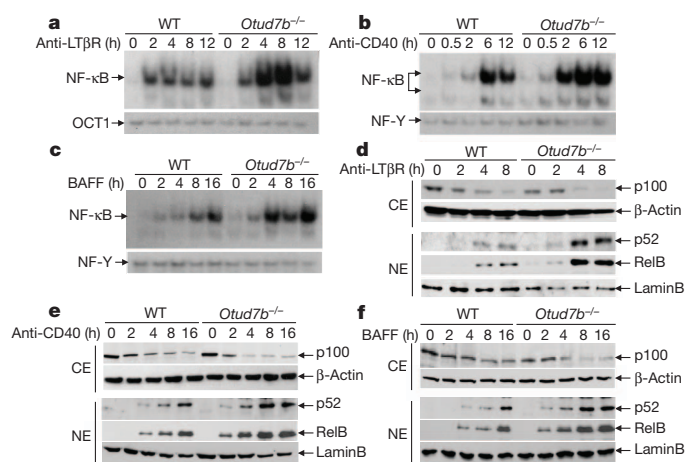


Figure 1 | OTUD7B negatively regulates the non-canonical NF- κ B pathway. **a–c**, Electrophoretic mobility-shift assay of NF- κ B and control DNA-binding factors (OCT1, NF-Y) using nuclear extracts isolated from wild-type (WT) and *Otud7b*^{−/−} primary MEFs (**a**) or B cells (**b**, **c**) stimulated with the indicated inducers. **d–f**, Immunoblot assays using cytoplasmic (CE) and nuclear (NE) extracts of wild-type and *Otud7b*^{−/−} primary MEFs (**d**) or B cells (**e**, **f**) that were stimulated as indicated. Data are representative of three independent experiments.

¹Department of Immunology, The University of Texas MD Anderson Cancer Center, 7455 Fannin Street, Box 902, Houston, Texas 77030, USA. ²Department of Pathology, University of Chicago, Chicago, Illinois 60637, USA. ³The University of Texas Graduate School of Biomedical Sciences, Houston, Texas 77030, USA.

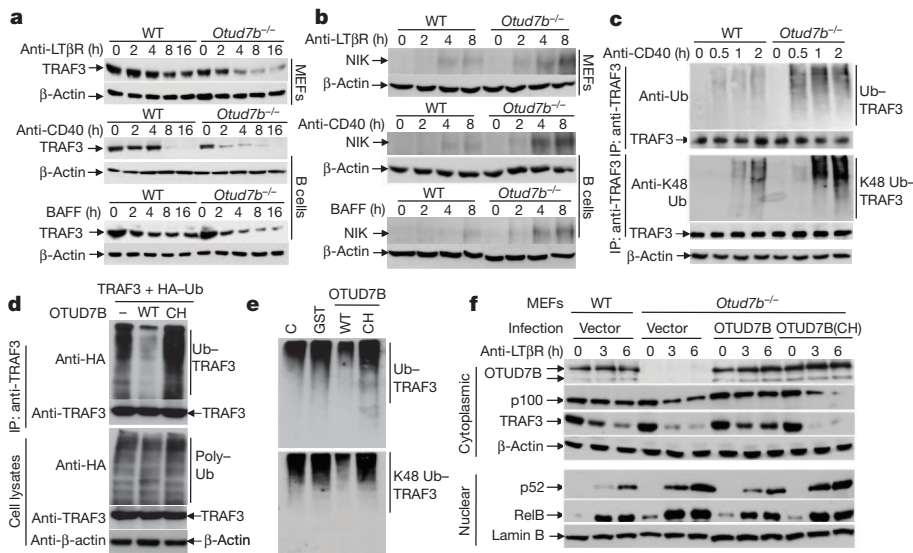


Figure 2 | OTUD7B negatively regulates TRAF3 degradation by affecting TRAF3 ubiquitination. **a, b**, Immunoblot analyses of TRAF3 and NIK in wild-type and *Otud7b*^{-/-} MEFs or B cells, stimulated as indicated. **c**, Endogenous TRAF3 was isolated by immunoprecipitation, under denaturing conditions, from wild-type or *Otud7b*^{-/-} B cells treated with anti-CD40 plus MG132. Total and K48 ubiquitination of TRAF3 were detected by immunoblot using anti-ubiquitin and anti-K48-ubiquitin antibodies, respectively. **d**, HEK293 cells were transfected with the indicated expression vectors and

subjected to TRAF3 ubiquitination (upper) and protein expression (lower) assays. **e**, TRAF3 was transfected with HA-ubiquitin in HEK293 cells. Purified TRAF3-ubiquitin conjugates were incubated with buffer control (C), GST, GST-OTUD7B, or GST-OTUD7B(CH) recombinant proteins followed by anti-HA immunoblot. **f**, Immunoblot analysis of subcellular extracts from wild-type or *Otud7b*^{-/-} MEFs, stably infected with vector, OTUD7B, or OTUD7B(CH) and stimulated as indicated. Data are representative of at least three independent experiments.

to the quality of the anti-K11 ubiquitin antibody or the dominant K48 ubiquitination of TRAF3 (ref. 7). However, K48 ubiquitination of TRAF3 was markedly enhanced in OTUD7B-deficient cells (Fig. 2c). Because OTUD7B also displays K48-specific DUB function in transfected cells¹⁶, it is conceivable that the specificity of OTUD7B may be regulated by cofactors *in vivo*, as implicated for A20¹⁰. The OTUD7B-mediated TRAF3 deubiquitination required its DUB activity, as a catalytically inactive OTUD7B mutant, C194S/H358R (CH), failed to deubiquitinate TRAF3 (Fig. 2d, e). Consistently, wild-type OTUD7B,

but not OTUD7B(CH), suppressed the activation of non-canonical NF-κB signalling (Fig. 2f).

In response to CD40 or LTβR stimulation, OTUD7B was rapidly recruited to TRAF3 (Fig. 3a), and this molecular interaction was coupled with the receptor recruitment of OTUD7B along with TRAF3, TRAF2 and c-IAP2 (Fig. 3b, c). This finding was intriguing, because TRAF3 ubiquitination is mediated by c-IAP following their recruitment to the stimulating receptors^{2,7}. Although OTUD7B deficiency did not affect the recruitment of TRAF2, TRAF3 and c-IAP (Fig. 3c), TRAF3

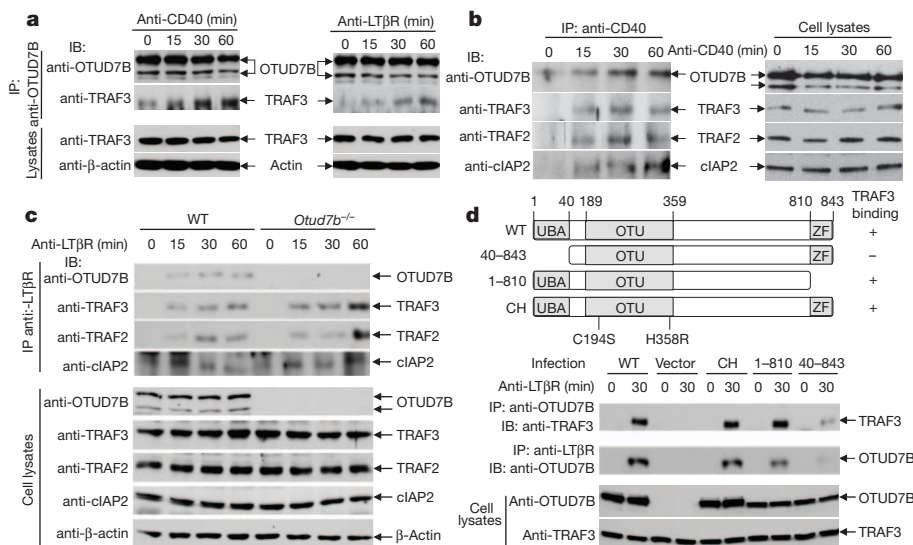


Figure 3 | OTUD7B inducibly interacts with TRAF3 and is recruited to the receptor complexes. **a**, M12 B cells (left) and wild-type MEFs (right) were stimulated and subjected to OTUD7B and TRAF3 co-immunoprecipitation (upper) and direct immunoblot (lower) assays. **b**, M12 hCD40 cells were stimulated with anti-human CD40. CD40 was isolated by immunoprecipitation, and its associated proteins were identified by immunoblot (left panel). Cell lysates were subjected to direct immunoblot (right panel). **c**, Wild-type and *Otud7b*^{-/-}

MEFs were stimulated and subjected to identification of LTβR-associated proteins (upper) and analysis of protein expression (lower). **d**, *Otud7b*^{-/-} MEFs, stably infected with wild-type OTUD7B or the indicated OTUD7B mutants, were stimulated with anti-LTβR and subjected to co-immunoprecipitation assays to detect OTUD7B-TRAF3 interaction (panel 1) and OTUD7B recruitment to LTβR (panel 2) or direct immunoblot assays (bottom two panels). Data are representative of two to three independent experiments.

knockdown blocked the receptor recruitment of OTUD7B (Supplementary Fig. 14). The ubiquitin-association (UBA) domain, but not the zinc finger (ZF) domain, of OTUD7B was required for its association with TRAF3 and receptor recruitment (Fig. 3d). The UBA domain of OTUD7B was also indispensable for negatively regulating non-canonical NF- κ B signalling (Supplementary Fig. 15). These findings indicate that OTUD7B engages TRAF3 in response to cellular stimuli and, thereby, is recruited to the receptor complex, where it may mediate TRAF3 deubiquitination.

Our finding that OTUD7B controls signal-induced, but not steady-state, activation of non-canonical NF- κ B suggests specific roles for this DUB in the regulation of immune function. *Otud7b*^{-/-} mice did not show obvious structural abnormalities of secondary lymphoid organs or the thymus, although the medullary thymic epithelial cells of *Otud7b*^{-/-} mice had elevated expression of the autoimmune regulator (AIRE) and several other related genes, indicating a higher level of maturation (data not shown). Furthermore, *Otud7b*^{-/-} mice displayed a remarkable phenotype in the mucosal tertiary lymphoid structures, characterized by a marked increase in the size and number of B-cell-containing colonic patches (CLPs) (Fig. 4a and Supplementary

Fig. 16). Consistently, *Otud7b*^{-/-} mice had a significant increase in faecal IgA concentration compared to wild-type mice (Fig. 4b). Intestine is an organ that maintains dynamic host–microbiota homeostasis¹⁷, and unlike the secondary lymphoid organs, the development of which is programmed in the fetus, the formation and maintenance of intestinal tertiary lymphoid tissues occur in adults and involve microbial triggers and LT β R signalling^{18–21}. Indeed, injection of *Otud7b*^{-/-} mice with a LT β R fusion protein (LT β R-Ig), known to block the interaction of LT β R with its ligands¹⁸, suppressed the lymphoid hyperplasia (Fig. 4c and Supplementary Fig. 17). Consistent with the involvement of chemokines in LT β R-mediated induction of lymphoid follicles, the intestines of *Otud7b*^{-/-} mice had elevated expression of two major chemokines, CXCL12 and CXCL13, and the cell adhesion molecule MADCAM1, which was suppressed upon injection of LT β R-Ig (Fig. 4d). The LT β R-mediated induction of these genes was also enhanced in the *Otud7b*^{-/-} MEFs (Fig. 4e).

The results described above prompted us to examine whether OTUD7B negatively regulates mucosal immunity against infections. Indeed, the *Otud7b*^{-/-} mice were considerably more resistant to a well-defined intestinal bacterial pathogen, *C. rodentium*²², as shown by

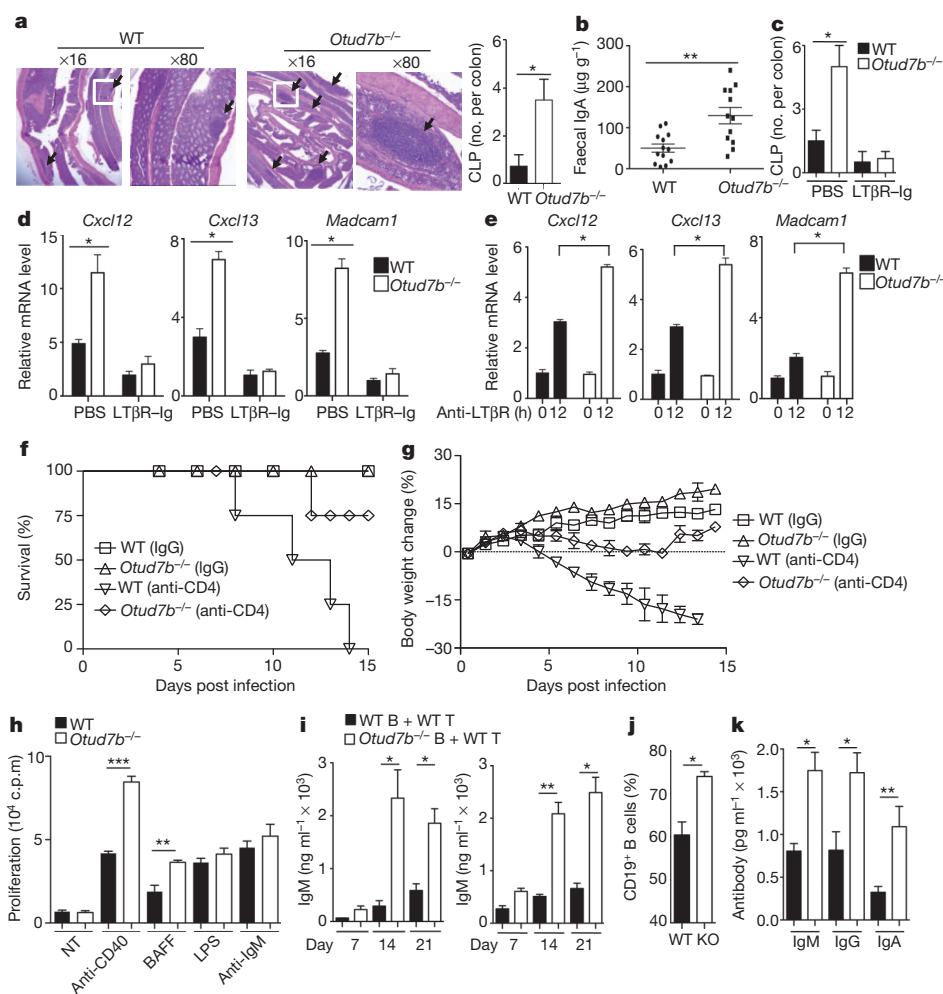


Figure 4 | OTUD7B negatively regulates intestinal lymphoid homeostasis, antibacterial immunity and B-cell responses. **a**, Haematoxylin and eosin staining (left) and summary graph (right) of wild-type and *Otud7b*^{-/-} colon ($n = 4$). Black arrows point to CLPs, and the boxed area is enlarged in the $\times 80$ pictures. **b**, ELISA of IgA in the faeces of wild-type and *Otud7b*^{-/-} mice ($\mu\text{g IgA}$ per gram of faeces, $n = 13$). **c**, Summary of CLP numbers in wild-type and *Otud7b*^{-/-} mice ($n = 4$) treated with PBS or LT β R-Ig. **d**, **e**, qPCR assays using RNAs prepared from colonic tissues (**d**) and anti-LT β R-stimulated MEFs (**e**). **f**, **g**, Survival (**f**) and body weight (**g**) of *C. rodentium*-infected wild-type or *Otud7b*^{-/-} mice that were also injected with a control IgG or an anti-CD4

antibody on day 0, 5 and 10 ($n = 4$). **h**, Proliferation assays of splenic B cells cultured without (NT) or with the indicated inducers *in vitro*. **i**, ELISA of serum 4-hydroxy-3-nitrophenyl (NP)-specific antibodies in NP-KLH-immunized *Rag1*^{-/-} mice adoptively transferred with wild-type or *Otud7b*^{-/-} B cells plus wild-type T cells. **j**, **k**, Mice were injected with BAFF (30 mg kg⁻¹) intravenously every other day for 2 weeks and killed on day 15 for flow cytometric analysis of B-cell frequency in splenocytes (**j**) and ELISA of the serum antibodies (**k**) ($n = 4$). Data are representative of two–three independent experiments. Bar graphs are presented as mean \pm s.d. values. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

the significantly reduced bacterial load in organs and faeces and the ameliorated crypt hyperplasia in the colon (Supplementary Fig. 18). To compare the *C. rodentium*-induced lethality in the wild-type and *Otud7b*^{-/-} mice, we took advantage of the recent finding that depletion of CD4 T cells in wild-type mice renders the mice sensitive to *C. rodentium*-induced lethality (Y.-X.F. *et al.*, unpublished data). Under these conditions, all of the wild-type mice died of *C. rodentium* infection by day 14, whereas 75% of the *Otud7b*^{-/-} mice survived infection (Fig. 4f). Consistently, *Otud7b*^{-/-} mice also experienced much milder body weight loss and had a significantly reduced faecal bacterial load compared to wild-type mice (Fig. 4g and Supplementary Fig. 19). Collectively, these findings indicate the possibility that inhibition of OTUD7B may enhance mucosal immunity against *C. rodentium* infection.

Deregulated non-canonical NF-κB activation in B cells, as seen in B-cell-conditional *Traf3*^{-/-} mice, is associated with B-cell hyperplasia and aberrant antibody production²³. Consistently, OTUD7B-deficient B cells displayed elevated proliferative and survival ability when stimulated *in vitro* (Fig. 4h and Supplementary Fig. 20) and were hyper-responsive to antigen-stimulated antibody production when adoptively transferred into *Rag1*^{-/-} mice along with wild-type T cells (Fig. 4i). We found that *Otud7b*^{-/-} mice did not show abnormal B-cell homeostasis under non-transferred conditions (Supplementary Fig. 4). This was probably due to attenuated activation of the OTUD7B-deficient T cells (data not shown), a phenotype that has also been observed with TRAF3-deficient T cells^{24,25}. Notably, upon injection with low doses of recombinant BAFF, the *Otud7b*^{-/-} mice displayed overt B-cell hyperplasia (Fig. 4j), coupled with accumulation of various antibody isotypes (Fig. 4k), thus further indicating a negative role for OTUD7B in regulating B-cell responses.

Our findings establish OTUD7B as a pivotal negative regulator of the non-canonical NF-κB pathway. Although another DUB, DUBA, has been implicated in TRAF3 deubiquitination, DUBA regulates the K63 ubiquitination and non-degradative function of TRAF3 in the pattern-recognition receptor (PRR)-induced IFN-I signalling pathway²⁶. So far, OTUD7B is the only DUB known to regulate TRAF3 degradation and non-canonical NF-κB signalling. OTUD7B is unique in that it specifically regulates signal-induced non-canonical NF-κB activation. These unique functions of OTUD7B make it an attractive candidate to be exploited as a therapeutic target to boost mucosal immunity and treat diseases associated with the non-canonical NF-κB pathway.

METHODS SUMMARY

Mice, infection and LTβR-Ig treatment. *Otud7b*^{-/-} mice were generated by a conventional gene-targeting approach (Deltagen, Inc.) and backcrossed for four generations to the C57BL/6 background. Age-matched wild-type and *Otud7b*^{-/-} mice were orally infected with *C. rodentium* and then killed for analyses of bacterial burdens and colon histology. In some experiments, the mice were depleted of CD4 T cells by injection of an anti-CD4 antibody before the *C. rodentium* infection. For assessing the role of LTβR signalling in colonic patch formation, the mice were injected intraperitoneally with PBS or LTβR-Ig on day 0 and day 5 and killed on day 10 for colon histology analyses.

Cell culture and stimulation. HEK-293 cell culture and transfection, splenic B-cell purification, and the generation of BMDMs and MEFs were as described²⁷. The cells were stimulated with the indicated inducers and subjected to cell proliferation or immunoblot assays.

Statistical analysis. Two-tailed unpaired *t*-tests were performed using Prism software. *P* values less than 0.05 were considered significant, and the level of significance was indicated as **P* < 0.05, ***P* < 0.01 and ****P* < 0.001.

Full Methods and any associated references are available in the online version of the paper.

Received 8 February; accepted 3 December 2012.

Published online 20 January 2013.

- Dejardin, E. The alternative NF-κB pathway from biochemistry to biology: pitfalls and promises for future drug development. *Biochem. Pharmacol.* **72**, 1161–1179 (2006).
- Sun, S. C. The noncanonical NF-κB pathway. *Immunol. Rev.* **246**, 125–140 (2012).

- Razani, B., Reichardt, A. D. & Cheng, G. Non-canonical NF-κB signaling activation and regulation: principles and perspectives. *Immunol. Rev.* **244**, 44–54 (2011).
- Senftleben, U. *et al.* Activation of IKKα of a second, evolutionary conserved, NF-κB signaling pathway. *Science* **293**, 1495–1499 (2001).
- Xiao, G., Harhaj, E. W. & Sun, S. C. NF-κB-inducing kinase regulates the processing of NF-κB2 p100. *Mol. Cell* **7**, 401–409 (2001).
- Liao, G., Zhang, M., Harhaj, E. W. & Sun, S. C. Regulation of the NF-κB-inducing kinase by tumor necrosis factor receptor-associated factor 3-induced degradation. *J. Biol. Chem.* **279**, 26243–26250 (2004).
- Vallabhapurapu, S. *et al.* Nonredundant and complementary functions of TRAF2 and TRAF3 in a ubiquitination cascade that activates NIK-dependent alternative NF-κB signaling. *Nature Immunol.* **9**, 1364–1370 (2008).
- Zarnegar, B. J. *et al.* Noncanonical NF-κB activation requires coordinated assembly of a regulatory complex of the adaptors cIAP1, cIAP2, TRAF2 and TRAF3 and the kinase NIK. *Nature Immunol.* **9**, 1371–1378 (2008).
- Coornaert, B., Carpentier, I. & Beyaert, R. A20: central gatekeeper in inflammation and immunity. *J. Biol. Chem.* **284**, 8217–8221 (2009).
- Harhaj, E. W. & Dixit, V. M. Regulation of NF-κB by deubiquitinases. *Immunol. Rev.* **246**, 107–124 (2012).
- Evans, P. C. *et al.* Isolation and characterization of two novel A20-like proteins. *Biochem. J.* **357**, 617–623 (2001).
- Lee, E. G. *et al.* Failure to regulate TNF-induced NF-κB and cell death responses in A20-deficient mice. *Science* **289**, 2350–2354 (2000).
- Häcker, H. *et al.* Specificity in Toll-like receptor signalling through distinct effector functions of TRAF3 and TRAF6. *Nature* **439**, 204–207 (2006).
- Oganesyan, G. *et al.* Critical role of TRAF3 in the Toll-like receptor-dependent and -independent antiviral response. *Nature* **439**, 208–211 (2006).
- Bremm, A., Freund, S. M. & Komander, D. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature Struct. Mol. Biol.* **17**, 939–947 (2010).
- Enesa, K. *et al.* NF-κB suppression by the deubiquitinating enzyme Cezanne: a novel negative feedback loop in pro-inflammatory signaling. *J. Biol. Chem.* **283**, 7036–7045 (2008).
- Hooper, L. V. & Macpherson, A. J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nature Rev. Immunol.* **10**, 159–169 (2010).
- Dohi, T. *et al.* Elimination of colonic patches with lymphotoxin beta receptor-Ig prevents Th2 cell-type colitis. *J. Immunol.* **167**, 2781–2790 (2001).
- Gommerman, J. L. & Browning, J. L. Lymphotoxin/light, lymphoid microenvironments and autoimmune disease. *Nature Rev. Immunol.* **3**, 642–655 (2003).
- Lorenz, R. G., Chaplin, D. D., McDonald, K. G., McDonough, J. S. & Newberry, R. D. Isolated lymphoid follicle formation is inducible and dependent upon lymphotoxin-sufficient B lymphocytes, lymphotoxin β receptor, and TNF receptor I function. *J. Immunol.* **170**, 5475–5482 (2003).
- Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* **456**, 507–510 (2008).
- Wang, Y. *et al.* Lymphotoxin beta receptor signaling in intestinal epithelial cells orchestrates innate immune responses against mucosal bacterial infection. *Immunity* **32**, 403–413 (2010).
- Xie, P., Stunz, L. L., Larison, K. D., Yang, B. & Bishop, G. A. Tumor necrosis factor receptor-associated factor 3 is a critical regulator of B cell homeostasis in secondary lymphoid organs. *Immunity* **27**, 253–267 (2007).
- Xu, Y., Cheng, G. & Baltimore, D. Targeted disruption of TRAF3 leads to postnatal lethality and defective T-dependent immune responses. *Immunity* **5**, 407–415 (1996).
- Xie, P., Kraus, Z. J., Stunz, L. L., Liu, Y. & Bishop, G. A. TNF receptor-associated factor 3 is required for T cell-mediated immunity and TCR/CD28 signaling. *J. Immunol.* **186**, 143–155 (2011).
- Kayagaki, N. *et al.* DUBA: a deubiquitinase that regulates type I interferon production. *Science* **318**, 1628–1632 (2007).
- Chang, M., Jin, W. & Sun, S. C. Peli1 facilitates TRIF-dependent Toll-like receptor signaling and proinflammatory cytokine production. *Nature Immunol.* **10**, 1089–1095 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Z. Chen and P. Evans for expression vectors and J. Browning and Biogen for the anti-LTβR antibody. We also thank the personnel from the flow cytometry, DNA analysis, animal facility, and histology core facilities at The MD Anderson Cancer Center for technical assistance. This study was supported by grants from the National Institutes of Health (AI057555, AI064639 and GM84459 to S.-C.S.; CA137059 to T.Z.; T32CA009598 to G.C.B.) and the Sister Institution Network Fund of MD Anderson Cancer Center.

Author Contributions H.H. designed the study, performed experiments, analysed data and wrote part of the manuscript; G.C.B., J.-H.C., N.P.-O., J.J., A.Z., Y. X., X.C. and M.C. contributed to the performance of the experiments; Y.-X.F. contributed critical reagents; C.Z. and T.Z. were involved in the supervision of N.P.-O. and A.Z., respectively, and the discussion of results; and S.-C.S. designed the research and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.-C.S. (ssun@mdanderson.org).

METHODS

Mice. *Otud7b*^{-/-} mice (in the C57BL/6x129/sv genetic background) were generated by a conventional gene-targeting approach (Deltagen, Inc.) (Supplementary Fig. 1) and backcrossed for four generations to the C57BL/6 background. *Otud7b*^{+/-} heterozygous mice were bred to generate age-matched wild-type (+/+) and knockout (-/-) experimental mice. Mice were maintained in specific pathogen-free facility, and all animal experiments were conducted in accordance with protocols approved by the Institutional Animal Care and Use Committee of the University of Texas MD Anderson Cancer Center.

Plasmids, antibodies and reagents. HA-tagged OTUD7B (pHR6-Cezanne) was provided by P. Evans²⁸, and OTUD7B C194S/H358R (OTUD7B(CH)), harbouring point mutations in two conserved residues of the catalytic domain, was created by site-directed mutagenesis (Agilent Technologies). OTUD7B, OTUD7B(CH) and OTUD7B(1–810) and OTUD7B(40–843) were cloned into pCLXSN(GFP) vector²⁹ for MEF infections. OTUD7B and OTUD7B(CH) were cloned into pGEX-4T-1 bacterial expression vector for producing glutathione-S-transferase (GST) fusion proteins in *Escherichia coli*. The HA-tagged TRAF3 and ubiquitin vectors have been described previously⁶, and the HA-ubiquitin K48 (where all of the lysines, except for lysine 48, were substituted with alanines) was provided by Z. Chen.

Murine LTβR antibody (ACH6) was provided by J. Browning (Biogen). Anti-α-IAP2 (H-85) and anti-ubiquitin (P4D1) were from Santa Cruz Biotech, anti-K48 ubiquitin and anti-K11 ubiquitin were from Millipore, and anti-OTUD7B was from Proteintech. Other antibodies for cell stimulation, immunoblot and flow cytometry were as described^{6,27}. Recombinant LTβR-Ig was described previously³⁰. 5'ppp-dsRNA was from Invivogen, and other agents were as described²⁷.

Cell culture and stimulation. HEK-293 cell culture and transfection, splenic B-cell purification, and the generation of BMDMs and MEFs were as described²⁷. Thymic stromal cells were isolated as reported³¹. For B-cell proliferation assays, the cells were stimulated with anti-murine CD40 (500 ng ml⁻¹), BAFF (200 ng ml⁻¹), LPS (100 ng ml⁻¹) or anti-mouse IgM (10 μg ml⁻¹) for 40 h and then pulsed for 8 h with ³H-thymidine. Primary MEFs were immortalized by infection with a retroviral vector encoding the adenoviral E1A proto-protein (pCL-E1A). For OTUD7B reconstitution, the immortalized MEFs were infected with pCLXSN(GFP) encoding OTUD7B or its mutants and screened based on GFP expression.

For gene induction and signalling studies, BMDMs were stimulated with lipopolysaccharide (LPS) (100 ng ml⁻¹), poly(I:C) (10 μg ml⁻¹), R848 (2.5 μg ml⁻¹), or CpG ODN1668 (1 μg ml⁻¹). B cells were stimulated with anti-murine CD40 (500 ng ml⁻¹), BAFF (200–400 ng ml⁻¹), LPS (5 μg ml⁻¹), or anti-IgM (10 μg ml⁻¹), and MEFs were stimulated with anti-LTβR (500 ng ml⁻¹), LPS (1 μg ml⁻¹), TNF-α (50 ng ml⁻¹), IL-1β (10 ng ml⁻¹), or lipofectamine-transfected poly(I:C) (2.5 μg ml⁻¹) or dsRNA (1 μg ml⁻¹) (for stimulation of RIG-I).

Receptor recruitment assays. The recruitment of OTUD7B and other signalling molecules to CD40 was detected as previously described³². Briefly, M12 cells stably expressing human CD40 (M12-hCD40) (2 × 10⁷) were stimulated with anti-hCD40 (in 1 ml of growth medium) and lysed in 600 μl of a lysis buffer. The CD40 complexes were precipitated by protein G-agarose and analysed by immunoblotting. The recruitment of proteins to LTβR was detected in a similar way, except for the use of MEFs stimulated with anti-LTβR antibody²².

Ubiquitination and deubiquitination assays. TRAF3 was isolated by immunoprecipitation under denaturing conditions²⁷ to inactivate DUBs and disrupt protein complexes, and the ubiquitinated TRAF3 was detected by immunoblotting using pan-ubiquitin or chain-specific ubiquitin antibodies. For transfection models, TRAF3 was transfected into HEK293 cells along with HA-ubiquitin or HA-ubiquitin K48. Following TRAF3-denaturing immunoprecipitation, the TRAF3-ubiquitin conjugates were detected by immunoblotting using anti-HA. In deubiquitination assays, the isolated TRAF3-ubiquitin conjugates were incubated with purified GST-OTUD7B or GST-OTUD7B(CH) recombinant proteins in a deubiquitination buffer³³ for 16 h and then subjected to immunoblotting using anti-HA.

Flow cytometry. Cell suspensions were subjected to flow cytometry analyses using a LSRII flow cytometer (BD Biosciences), as described previously³⁴. The data were analysed using FlowJo software.

Real-time quantitative reverse PCR (qPCR). RNA preparation and qPCR assays were as described²⁷ using gene-specific primers listed in Supplementary Table 1.

B-cell adoptive transfer, immunization and ELISA. B220⁺ B cells and CD90.2⁺ T cells were isolated from the splenocytes of wild-type or *Otud7b*^{-/-} mice using magnetic beads (Miltenyi Biotec). The isolated cells were >95% pure, as determined by flow cytometry. Wild-type T cells (5 × 10⁶) were mixed with either wild-type or *Otud7b*^{-/-} B cells (5 × 10⁶) and then injected via a tail vein into *Rag1*^{-/-} mice. After 16 h, the recipient mice were subjected to immunization with NP-KLH and the sera was collected on day 7, day 14 and day 21. ELISA was used to detect the different isotopes of antibodies (Southern Biotech). For measuring faecal IgA concentration, faeces were taken from age-matched wild-type and *Otud7b*^{-/-} mice, weighted, and homogenized in PBS. IgA level was detected by ELISA.

LTβR-Ig treatment. To block the LTβR signalling, wild-type or *Otud7b*^{-/-} mice were injected intraperitoneally with 100 μl of PBS or LTβR-Ig (1 μg μl⁻¹) on day 0 and day 5 and killed on day 10. Colons were collected for histological examination.

Electrophoretic mobility-shift assay (EMSA). Nuclear extracts were prepared from indicated cells and subjected to EMSA using the following ³²P-radiolabelled oligonucleotide probes: NF-κB, 5'-CAACGGCAGGGGAATCCCCCTCTCCTT-3'; NF-Y, 5'-AAGAGATTAACCAATCACGTACGGTCT-3'; OCT1, 5'-TGTC GAATGCAAACTCACTAGAA-3'.

C. rodentium infection. *C. rodentium* strain DBS100 (ATCC 51459) was cultured by shaking at 37 °C overnight in LB broth. The concentration of bacteria was assessed by measuring optical density at 600 nm and confirmed by plating serially diluted bacterial cultures for colony-forming unit (c.f.u.) determination. Mice were orally injected with 2–4.5 × 10⁹ *C. rodentium* in a total volume of 200 μl per mouse. For CD4 T-cell depletion, the mice were also injected intraperitoneally with either a control rat IgG or a rat anti-mouse CD4 antibody (GK1.5, 50 μg per mouse) on day 0, 5 and 10. Mice were allowed access to food and water after the inoculation and killed at the indicated times after infection. Colons were isolated and fixed for haematoxylin and eosin staining to evaluate tissue pathology. Spleen, liver and faeces were collected, weighed and homogenized. The homogenates were serially diluted and plated on MacConkey agar plates, and the plates were incubated at 37 °C for 24 h followed by counting the *C. rodentium* colonies (pink colonies).

CLP staining. Whole-mount immunofluorescence staining of colonic lymphoid follicles was performed essentially as described³⁵. Colons were collected, flushed and opened longitudinally along the mesenteric border. The tissues were incubated with shaking at room temperature for 60 min in Hank's balanced-salt solution containing 5 mM EDTA, fixed in 4% paraformaldehyde, and rinsed. After blocking nonspecific epitopes with a blocking buffer³⁵ containing 5% BSA, colons were stained with Alexa Fluor 594-conjugated anti-B220 antibody. Pictures were taken by Leica SP5 RS confocal microscope and analysed by SlideBook 5.0 software. The B220⁺ lymphoid follicles larger than 10,000 μm² were considered as CLPs.

Statistical analysis. Two-tailed unpaired *t*-tests were performed using Prism software. *P* values less than 0.05 were considered significant, and the level of significance was indicated as **P* < 0.05, ***P* < 0.01 and ****P* < 0.001.

28. Evans, P. C. *et al.* A novel type of deubiquitinating enzyme. *J. Biol. Chem.* **278**, 23180–23186 (2003).
29. Reiley, W., Zhang, M., Wu, X., Graner, E. & Sun, S.-C. Regulation of the deubiquitinating enzyme CYLD by IκB kinase γ-dependent phosphorylation. *Mol. Cell. Biol.* **25**, 3886–3895 (2005).
30. Anders, R. A., Subudhi, S. K., Wang, J., Pfeffer, K. & Fu, Y. X. Contribution of the lymphotoxin β receptor to liver regeneration. *J. Immunol.* **175**, 1295–1300 (2005).
31. Lomada, D., Liu, B., Coghlan, L., Hu, Y. & Richie, E. R. Thymus medulla formation and central tolerance are restored in IKKα^{-/-} mice that express an IKKα transgene in keratin 5⁺ thymic epithelial cells. *J. Immunol.* **178**, 829–837 (2007).
32. Morrison, M. D., Reiley, W., Zhang, M. & Sun, S. C. An atypical tumor necrosis factor (TNF) receptor-associated factor-binding motif of B cell-activating factor belonging to the TNF family (BAFF) receptor mediates induction of the noncanonical NF-κB signaling pathway. *J. Biol. Chem.* **280**, 10018–10024 (2005).
33. Hassink, G. C. *et al.* The ER-resident ubiquitin-specific protease 19 participates in the UPR and rescues ERAD substrates. *EMBO Rep.* **10**, 755–761 (2009).
34. Reiley, W. W. *et al.* Deubiquitinating enzyme CYLD negatively regulates the ubiquitin-dependent kinase Tak1 and prevents abnormal T cell responses. *J. Exp. Med.* **204**, 1475–1485 (2007).
35. Ota, N. *et al.* IL-22 bridges the lymphotoxin pathway with the maintenance of colonic lymphoid structures during infection with *Citrobacter rodentium*. *Nature Immunol.* **12**, 941–948 (2011).

Central role of E3 ubiquitin ligase MG53 in insulin resistance and metabolic disorders

Ruisheng Song^{1,2*}, Wei Peng^{1*}, Yan Zhang^{1*}, Fengxiang Lv¹, Hong-Kun Wu¹, Jiaojiao Guo¹, Yongxing Cao³, Yanbin Pi³, Xin Zhang³, Li Jin¹, Mao Zhang¹, Peng Jiang¹, Fenghua Liu¹, Shaoshuai Meng¹, Xiuqin Zhang¹, Ping Jiang¹, Chun-Mei Cao¹ & Rui-Ping Xiao^{1,4}

Insulin resistance is a fundamental pathogenic factor present in various metabolic disorders including obesity and type 2 diabetes¹. Although skeletal muscle accounts for 70–90% of insulin-stimulated glucose disposal^{2,3}, the mechanism underlying muscle insulin resistance is poorly understood. Here we show in mice that muscle-specific mitsugumin 53 (MG53; also called TRIM72) mediates the degradation of the insulin receptor and insulin receptor substrate 1 (IRS1), and when upregulated, causes metabolic syndrome featuring insulin resistance, obesity, hypertension and dyslipidaemia. MG53 expression is markedly elevated in models of insulin resistance, and MG53 overexpression suffices to trigger muscle insulin resistance and metabolic syndrome sequentially. Conversely, ablation of MG53 prevents diet-induced metabolic syndrome by preserving the insulin receptor, IRS1 and insulin signalling integrity. Mechanistically, MG53 acts as an E3 ligase targeting the insulin receptor and IRS1 for ubiquitin-dependent degradation, comprising a central mechanism controlling insulin signal strength in skeletal muscle. These findings define MG53 as a novel therapeutic target for treating metabolic disorders and associated cardiovascular complications.

A cluster of disorders known as metabolic syndrome, including insulin resistance, central obesity, dyslipidaemia and hypertension, is increasing at epidemic rate and has become one of the most serious threats to human health^{1,4}. Metabolic syndrome increases the risk of developing cardiovascular disease twofold, and the risk of type 2 diabetes fivefold⁵. Insulin resistance is a fundamental pathogenic factor shared by a myriad of metabolic disorders including metabolic syndrome, obesity and type 2 diabetes¹. Because skeletal muscle is responsible for 70–90% of insulin-stimulated glucose disposal^{2,3}, insulin resistance in skeletal muscle probably has a central role in the pathogenesis of metabolic syndrome and resultant type 2 diabetes⁶. Indeed, longitudinal studies have provided evidence that skeletal muscle insulin resistance is the earliest step in the pathogenesis of metabolic syndrome and type 2 diabetes^{7,8}. However, the mechanism underlying skeletal muscle insulin resistance is poorly understood.

MG53 is a recently identified tripartite motif-containing (TRIM) family protein and is specifically expressed in skeletal muscle and the heart^{9,10} (Supplementary Fig. 1a). Recent studies have demonstrated that MG53 participates in muscle membrane repair⁹ and cardiac ischaemic preconditioning¹⁰. A survey of MG53 expression in animal models for insulin resistance and metabolic disorders^{11–14} revealed that its abundance universally increases in high-fat diet (HFD)-induced obese mice, *db/db* diabetic mice, spontaneously hypertensive rats, and nonhuman primates with metabolic syndrome (Fig. 1a, b). The upregulation of MG53 was also confirmed in obese humans (Supplementary Fig. 2). These results provided the critical clue for a previously unappreciated link between MG53 and metabolic diseases.

To determine whether MG53 is required for the pathogenesis of metabolic syndrome, we used MG53-deficient (*MG53*^{−/−}) mice^{9,10} and their wild-type littermates (Supplementary Fig. 1b, c) to track,

from 3 weeks of age, changes in their body weight and metabolic parameters in response to a HFD (60% calories from fat) or chow. Compared to wild-type mice, *MG53*^{−/−} mice on chow showed no phenotypic difference in body weight, blood pressure, serum cholesterol and triglyceride levels from 3 to 38 weeks of age, except for a significant reduction in blood glucose concentration without changing the serum insulin level (measured at 38 weeks) (Fig. 1c–h).

Using dietary intervention, however, we found profound differences between *MG53*^{−/−} and wild-type mice. After 35 weeks on the HFD, with upregulated MG53 (Fig. 1a, b), wild-type mice developed metabolic syndrome featuring obesity and hypertension (Fig. 1c, d and Supplementary Fig. 3a–e), hyperglycaemia, hyperinsulinaemia, dyslipidaemia and hepatosteatosis (Fig. 1e–i). Notably, *MG53*^{−/−} mice were resistant to the HFD-induced metabolic disorder. After the HFD for 35 weeks, blood pressure, blood glucose and serum insulin and lipid (cholesterol and triglyceride) levels in *MG53*^{−/−} mice were largely comparable to those of wild-type or *MG53*^{−/−} mice on chow (Fig. 1d–h). MG53 deficiency also markedly attenuated HFD-induced

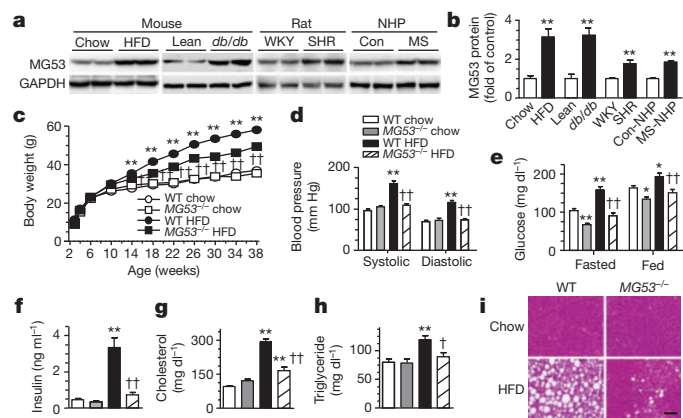


Figure 1 | MG53 ablation protects mice against diet-induced metabolic syndrome. **a, b**, Representative western blots and averaged data showing upregulation of MG53 in skeletal muscle from rodent and nonhuman primate (NHP) models of insulin resistance and metabolic disorders, versus skeletal muscle from their respective age- and gender-matched controls. Data are normalized to GAPDH (data are mean \pm s.e.m.; $n = 6$ chow, HFD; $n = 4$ for all of the other groups; $^{**}P < 0.01$ versus respective control). Con, control; MS, metabolic syndrome monkeys; SHR, spontaneously hypertensive rats; WKY, Wistar Kyoto rats. **c**, Body weights of wild-type (WT) and *MG53*^{−/−} mice on chow or the HFD for 35 weeks ($n = 30$ WT chow, $n = 19$ *MG53*^{−/−} chow, $n = 41$ WT HFD, $n = 27$ *MG53*^{−/−} HFD). **d–i**, Systolic and diastolic blood pressure (**d**) ($n = 8$), blood glucose levels in the fasted and fed state (**e**) ($n = 10$), serum insulin (**f**) ($n = 10$), cholesterol (**g**) ($n = 14$) and triglyceride levels (**h**) ($n = 23$), and haematoxylin and eosin staining of liver (**i**) of wild-type and *MG53*^{−/−} mice on chow or the HFD for 35 weeks. Scale bar in **i**, 25 μ m. Data are mean \pm s.e.m. for panels **c–h** ($^{*}P < 0.05$, $^{**}P < 0.01$ versus WT chow; $^{\dagger}P < 0.05$, $^{\dagger\dagger}P < 0.01$ versus WT HFD).

¹Institute of Molecular Medicine, State Key Laboratory of Biomembrane and Membrane Biotechnology, Peking University, Beijing 100871, China. ²Institute of Cardiovascular Sciences, Health Science Center, Peking University, Beijing 100083, China. ³Institute of Sports Medicine, Peking University Third Hospital, Beijing 100191, China. ⁴Center for Life Sciences, Peking University, Beijing 100871, China. *These authors contributed equally to this work.

obesity (Fig. 1c and Supplementary Fig. 3a, b), hepatosteatosis (Fig. 1i), lipid accumulation in skeletal muscle, adipocyte hypertrophy, and the increases in both white and brown fat weights (Supplementary Fig. 3c–e). In addition, the HFD treatment led to a significant decrease of daily energy expenditure (in both light and dark) in wild-type but not *MG53*^{-/-} mice (Supplementary Fig. 4d–f), whereas there was no phenotypic difference between these mice in their daily food intake or core body temperature (Supplementary Fig. 4a–c). Thus, deficiency of the muscle-specific protein MG53 exerts robust protective effects against HFD-triggered metabolic disorders and cardiovascular complications such as hypertension, underscoring a causal relationship between muscle MG53 upregulation and HFD-induced metabolic syndrome.

To track whole-body insulin sensitivity in relation to the development of metabolic syndrome, we performed glucose tolerance tests (GTTs) and insulin tolerance tests (ITTs) at various time points after dietary intervention. The HFD led to glucose intolerance and insulin resistance in wild-type mice at as early as 7 weeks (Fig. 2 and Supplementary Fig. 5), that is, before the onset of obesity at 11 weeks (Fig. 1c). As the HFD treatment was continued, both glucose intolerance and insulin resistance were progressively exacerbated in wild-type mice. After 35 weeks, the HFD caused compensatory hypertrophy of pancreatic islets (Fig. 2c and Supplementary Fig. 6), a sevenfold increase in basal serum insulin concentration (Fig. 2d), and failure of glucose to stimulate insulin secretion (Fig. 2e). Notably, *MG53*^{-/-} mice on the HFD showed none of these phenotypes, maintaining normal blood glucose and insulin levels even after 35 weeks of the HFD (Fig. 2a, b, d and Supplementary Fig. 5). Changes in pancreatic morphology and the insulin secretion response were also markedly ameliorated in *MG53*^{-/-} mice (Fig. 2c–e and Supplementary Fig. 6). Thus, MG53 ablation protects mice against HFD-induced insulin resistance and the sequelae of metabolic disorders, indicating that MG53 is required for HFD-induced insulin resistance and metabolic syndrome.

To determine further whether upregulation of MG53 is sufficient to initiate the pathogenesis of metabolic disorders, we generated transgenic mice overexpressing MG53 (*MG53*^{Tg}) (Supplementary Fig. 7a–c). MG53 protein levels were increased by 2.6 ± 0.3 -fold in skeletal muscle and 2.6 ± 0.7 -fold in the heart at 38 weeks of age (Supplementary Fig. 7c). However, there was no detectable expression of MG53 in non-muscle

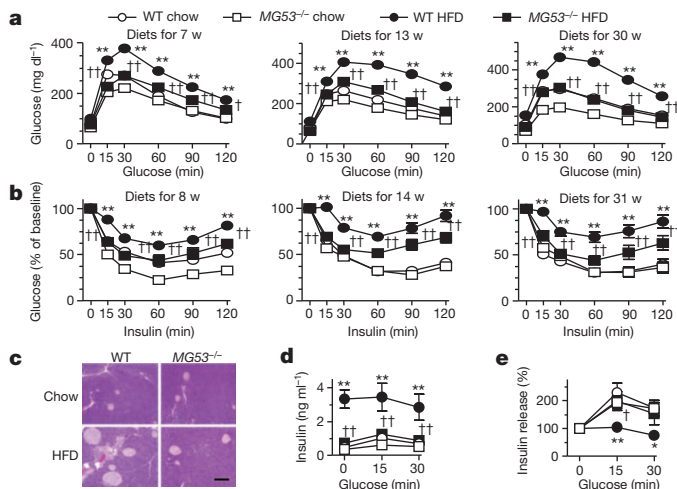


Figure 2 | MG53 ablation blocks diet-induced systemic insulin resistance. a, b, GTTs and ITTs were performed in wild-type and *MG53*^{-/-} mice on chow or the HFD at indicated time points ($n = 12$ WT chow, *MG53*^{-/-} HFD; $n = 23$ WT HFD; $n = 10$ *MG53*^{-/-} chow). c, Haematoxylin and eosin staining of the pancreas. Scale bar, 100 μ m. d, e, Glucose (2 g kg⁻¹, intraperitoneally) stimulated changes in serum insulin concentrations (d) or their relative changes as percentage of the respective baselines (e) ($n = 6$). In c–e, wild-type and *MG53*^{-/-} mice were on chow or the HFD for 35 weeks. Data are mean \pm s.e.m. (* $P < 0.05$, ** $P < 0.01$ versus WT chow; † $P < 0.05$, †† $P < 0.01$ *MG53*^{-/-} HFD versus WT HFD).

tissues (lung, brain, hypothalamus, liver, intestine, kidney, visceral fat and testis) in *MG53*^{Tg} mice, attesting that MG53 expression is tightly regulated in a muscle-specific manner (Supplementary Fig. 7c). Even in the absence of dietary intervention, *MG53*^{Tg} mice at 38 weeks of age were obese and hypertensive, along with dyslipidaemia, hyperinsulinaemia and increased fasted blood glucose levels (Fig. 3a–e). Energy expenditure during light and dark was significantly lower, but daily food intake was unchanged in *MG53*^{Tg} mice relative to wild-type littermates (Supplementary Fig. 8). Notably, GTTs and ITTs revealed severe impairments in glucose metabolism and insulin sensitivity in *MG53*^{Tg} mice (Fig. 3f, g), which were accompanied by pancreatic islet hypertrophy and failure of glucose-stimulated insulin secretion (Fig. 3h, i). Anatomical and histological data further documented that *MG53*^{Tg} mice had central obesity, hepatosteatosis, enlargement of adipocytes, and lipid accumulation in skeletal muscle (Supplementary Fig. 7d–g). These data indicate that upregulation of MG53 is both necessary and sufficient to trigger insulin resistance that further develops into metabolic syndrome.

Upon insulin stimulation, the intrinsic tyrosine kinase of the insulin receptor leads to receptor autophosphorylation at tyrosine residues¹⁵. Subsequent recruitment and phosphorylation of the insulin receptor substrates such as IRS1 and IRS2 is the pivotal event which, in turn, activates the downstream phosphatidylinositol-3-OH kinase (PI(3)K)–Akt–GSK3- β signalling pathway to regulate glucose homeostasis in skeletal muscle¹⁶. To decipher the molecular mechanism underlying the MG53-mediated insulin resistance, we investigated possible MG53-mediated regulation of the insulin receptor–IRS1–PI(3)K–Akt–GSK3- β signalling cascade. In the *MG53*^{Tg} model of insulin resistance and metabolic disorders, insulin-stimulated tyrosine phosphorylation of the insulin receptor (β subunit) and IRS1 and Ser 473 phosphorylation of Akt in skeletal muscle were abrogated (Fig. 4a, b). The protein levels of the insulin receptor and IRS1 were markedly reduced (Fig. 4a, b), whereas their messenger RNA levels remained intact in skeletal muscle from these mice (Supplementary Fig. 9a, b). To distinguish the direct effects of MG53 overexpression from adaptive responses, we performed adenoviral gene transfer of MG53 in cultured C2C12 myotubes. A 3.5 ± 0.2 -fold increase in MG53 over baseline levels led to marked reductions in the insulin receptor and IRS1 protein levels (Supplementary Fig. 10) but not their mRNA levels (Supplementary Fig. 9c, d).

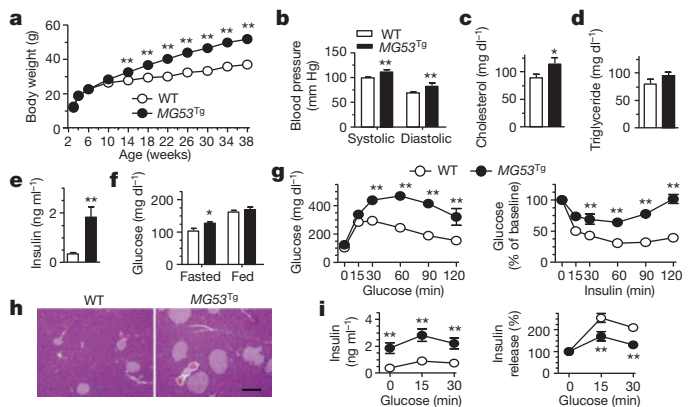


Figure 3 | Overexpression of MG53 triggers systemic insulin resistance and metabolic syndrome. a, Body weights of wild-type and *MG53*^{Tg} mice. b–f, Systolic and diastolic blood pressure (b), serum cholesterol (c), triglyceride (d), serum insulin (e) and blood glucose levels in the fasted and fed state (f) of wild-type and *MG53*^{Tg} mice at the age of 38 weeks. g, GTTs and ITTs in wild-type and *MG53*^{Tg} mice at the age of 30 and 31 weeks, respectively. h, Haematoxylin and eosin staining of the pancreas from wild-type and *MG53*^{Tg} mice at the age of 38 weeks. Scale bar, 50 μ m. i, Insulin concentrations or their relative changes as a percentage of respective baselines in response to glucose (2 g kg⁻¹, intraperitoneally) in wild-type and *MG53*^{Tg} mice at the age of 38 weeks. Data are mean \pm s.e.m. ($n = 10$ for panel a; $n = 8$ for panels b–g, i; * $P < 0.05$, ** $P < 0.01$ versus the corresponding wild-type group).

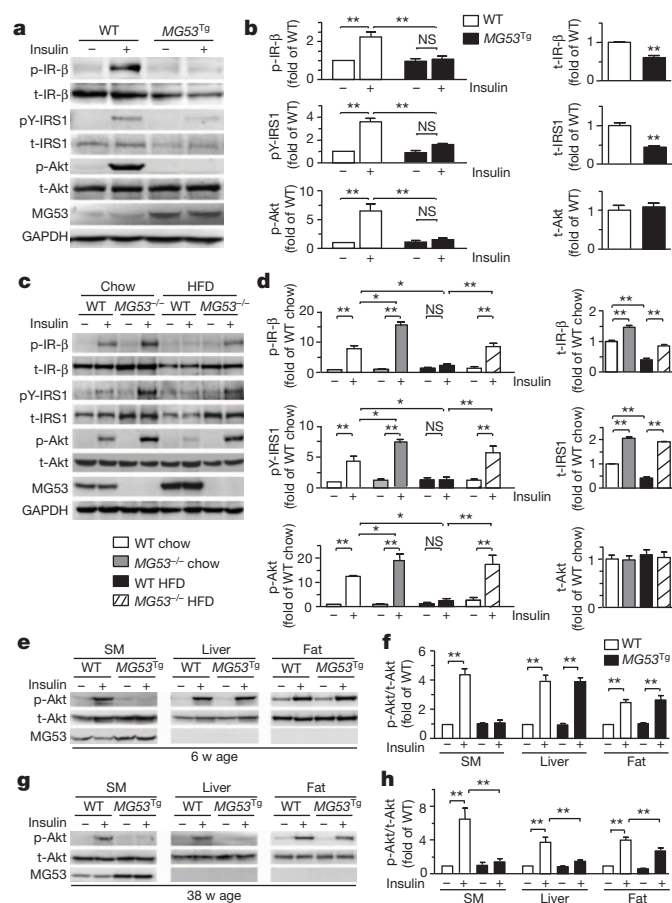


Figure 4 | Regulation of muscle insulin signalling by MG53.

a–d, Representative western blots and averaged data showing insulin-induced tyrosine phosphorylation of insulin receptor- β subunit (IR- β) and IRS1, serine phosphorylation of Akt and their total protein levels in skeletal muscle from wild-type and $MG53^{Tg}$ mice at the age of 38 weeks (**a**, **b**) or $MG53^{-/-}$ mice and their wild-type littermates on chow or the HFD for 35 weeks (**c**, **d**) ($n = 8$ for all groups). In **b** and **d**, phosphorylated and total proteins are normalized to GAPDH and presented as fold of their respective wild-type baselines. NS, not significant; p-, phosphorylated; t-, total. **e–h**, Typical western blots and statistical data showing insulin-induced phosphorylation of Akt in skeletal muscle (SM), liver and visceral fat (Fat) from wild-type and $MG53^{Tg}$ mice at the age of 6 weeks (**e**, **f**) or 38 weeks (**g**, **h**) ($n = 4$ for all groups). For insulin stimulation, animals were pre-treated with insulin (1 U kg^{-1} , intraperitoneally) for 15 min. Data are mean \pm s.e.m. (* $P < 0.05$, ** $P < 0.01$ as indicated).

MG53 overexpression also blocked the insulin-induced activation of the insulin receptor, IRS1, Akt and GSK3- β (Supplementary Fig. 10), indicating that upregulation of MG53 at the cellular level recapitulates the salient features of $MG53^{Tg}$ mice. Thus, our *in vivo* and *in vitro* data indicate that elevated MG53 expression is sufficient to cause simultaneous downregulation of the insulin receptor and IRS1, thus suppressing insulin signal transduction at two crucial steps.

Similar to MG53 overexpression, the HFD, which elevated MG53 expression level (Fig. 1a), profoundly suppressed skeletal muscle insulin signalling in wild-type mice (Fig. 4c, d and Supplementary Fig. 11) and simultaneously reduced insulin receptor and IRS1 protein levels without altering their mRNA levels (Fig. 4c, d and Supplementary Fig. 9e, f). Moreover, post-transcriptional downregulation of both the insulin receptor and IRS1, accompanied by MG53 upregulation, is a common feature shared by all of the rodent models of insulin resistance used in this study (Supplementary Figs 12 and 13), as is the case in animal models and humans with obesity or type 2 diabetes^{17–19}.

In more stringent tests, we showed that MG53 ablation enabled mice to maintain insulin receptor and IRS1 integrity as well as whole-body insulin sensitivity, even under the metabolic stress induced by the

HFD. Specifically, the HFD failed to decrease insulin receptor and IRS1 abundance in $MG53^{-/-}$ mice (Fig. 4c, d). In fact, genetic ablation of MG53 led to an evident accumulation of the insulin receptor and IRS1 in skeletal muscle, perhaps contributing to the decrease in blood glucose level in $MG53^{-/-}$ mice relative to wild-type littermates (Fig. 1e). Concomitantly, the insulin-induced phosphorylation of the insulin receptor, IRS1, Akt and GSK3- β was significantly greater in $MG53^{-/-}$ than that in wild-type mice on chow (Fig. 4c, d and Supplementary Fig. 11). Hence, MG53 upregulation seems to be indispensable for the HFD- and metabolic-disease-associated downregulation of the insulin receptor and IRS1 in skeletal muscle.

Next, we sought to determine how MG53-mediated, muscle-specific insulin resistance develops into whole-body metabolic disorders. For this purpose, we tracked the onset of various facets of metabolic disorders in multiple organs in $MG53^{Tg}$ mice and HFD-treated wild-type mice. The present results showed that skeletal muscle insulin resistance induced by overexpression of MG53 and the HFD preceded the development of whole-body metabolic disorders including obesity and multi-organ insulin resistance, not the other way round. In $MG53^{Tg}$ mice at 6 weeks of age, with body weight comparable to wild-type controls (Fig. 3a), insulin signalling was clearly impaired in skeletal muscle, but not liver and visceral fat tissues (Fig. 4e, f). Likewise, the HFD in wild-type mice led to significant MG53 upregulation and a muscle-specific insulin signalling defect at as early as 1 week of dietary intervention (Supplementary Fig. 14a, b), before the manifestation of systemic glucose intolerance and insulin intolerance at 7 weeks (Fig. 2a, b) and appreciable obesity at 11 weeks on the HFD (Fig. 1c). Later, liver and fat tissues developed full-fledged insulin resistance in both $MG53^{Tg}$ mice on chow (at 38 weeks of age, Fig. 4g, h) and wild-type mice on the HFD (35 weeks of dietary intervention, Supplementary Fig. 14c, d). Together with the profound protective effects of MG53 ablation on metabolic disorders, these data strongly suggest that MG53-mediated muscle insulin resistance is a major factor in the instigation of the pathogenic process. In line with the present findings, previous studies on humans and animal models have also suggested an essential role of reduced muscle insulin signalling in the pathogenesis of obesity and glucose intolerance^{7,8,20,21}, albeit not without controversy^{22,23}.

Because IRS1 is a nodal point shared by insulin receptor- and insulin-like growth factor-1 (IGF-I)-receptor-mediated signalling pathways, both converging on the IRS1–Akt cascade, we investigated the effects of MG53 on muscle IGF-I signalling in addition to insulin signalling. By comparing the dose–response of insulin- or IGF-I-induced tyrosine phosphorylation of IRS1 in skeletal muscle from wild-type and MG53-deficient mice, we found that MG53 preferentially targets IRS1 in the setting of insulin signalling. Specifically, whereas MG53 ablation augmented the whole dose–response of tyrosine phosphorylation of IRS1 by insulin stimulation, it potentiated the effect of IGF-I only at high (1.0 and 10 mg kg^{-1} , intravenously) but not low IGF-I concentrations (0.1 mg kg^{-1} , intravenously) (Supplementary Fig. 15). If MG53 selectively targets insulin-receptor-mediated IRS1 signalling, the apparent regulation of IGF-I receptor signalling by MG53 might be attributable to heterodimerization of the insulin receptor and IGF-I receptor or cross-activation of insulin receptor by high dosage of IGF-I.

Because MG53 contains a canonical E3 ligase RING finger domain at the amino terminus⁹, we proposed that it may function as a muscle-specific E3 ligase targeting the insulin receptor and IRS1 for ubiquitin-dependent degradation. Multiple lines of evidence support this hypothesis. First, co-immunoprecipitation revealed a physical interaction of endogenous MG53 with the insulin receptor and IRS1 in skeletal muscle in the presence or absence of insulin stimulation (Fig. 5a), and of ectopically expressed MG53 with the insulin receptor and IRS1 in HEK 293 cells (Supplementary Fig. 16). Second, overexpression of MG53 abundantly increased the ubiquitination of both the insulin receptor and IRS1 in skeletal muscle from $MG53^{Tg}$ mice (Fig. 5b, c). Furthermore, the HFD profoundly augmented insulin receptor and IRS1 ubiquitination in skeletal muscle *in vivo* in wild-type,

but not *MG53^{-/-}* mice (Fig. 5d, e). The present findings demonstrate that MG53 E3 ligase facilitates ubiquitin-dependent degradation of both the insulin receptor and IRS1 in skeletal muscle, although several other E3 ligases have been previously implicated in IRS1 turnover in certain cell culture systems and tissues with FBXO40 being muscle specific^{24–28}. In

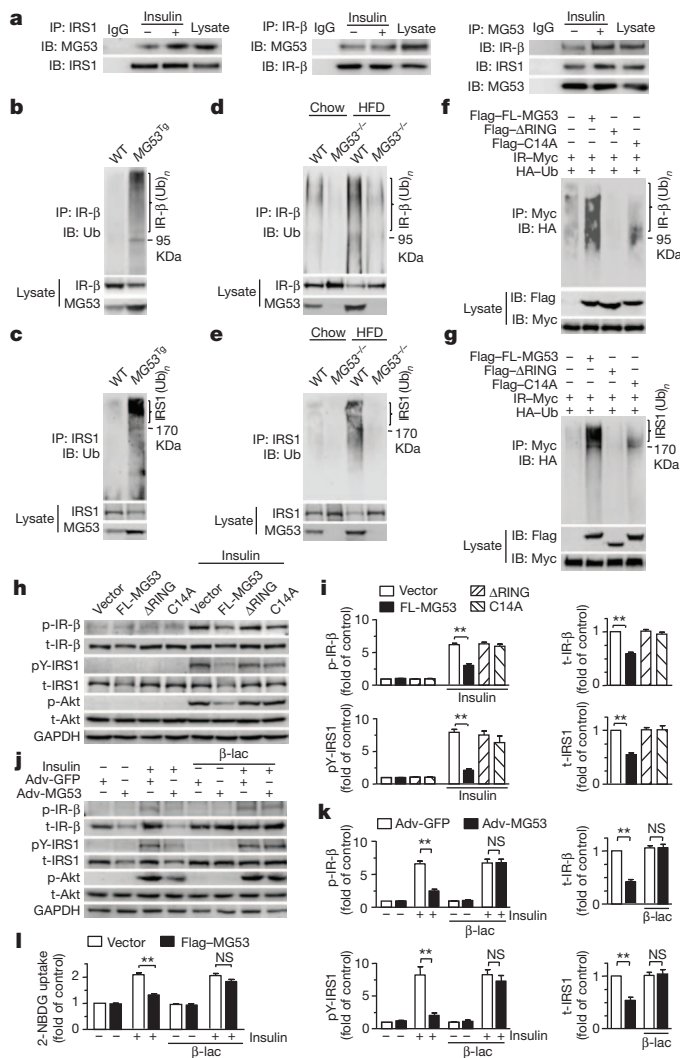


Figure 5 | MG53 E3 ligase activates ubiquitination of the insulin receptor and IRS1. **a**, Co-immunoprecipitation of MG53 with insulin receptor- β and IRS1 in mouse skeletal muscle with or without insulin pre-treatment (1 U kg^{-1} , intraperitoneally, for 15 min). The input represents 1% of the whole tissue lysate used for each immunoprecipitation. **b**, **c**, Ubiquitination of insulin receptor- β (**b**) and IRS1 (**c**) in skeletal muscle from wild-type and *MG53^{Tg}* mice at 38 weeks of age. **d**, **e**, Ubiquitination of insulin receptor- β (**d**) and IRS1 (**e**) in the skeletal muscles from wild-type and *MG53^{-/-}* mice on chow or the HFD for 35 weeks. **f**, **g**, Ubiquitination of insulin receptor- β (**f**) and IRS1 (**g**) in C2C12 myotubes expressing full-length MG53 (FL-MG53), or the RING domain deletion (Δ RING) or C14A mutant. **h**, **i**, Representative western blots (**h**) and averaged data (**i**) showing insulin-induced tyrosine phosphorylation of insulin receptor- β and IRS1 and serine phosphorylation of Akt in C2C12 myotubes expressing full-length MG53 (FL-MG53), or the Δ RING or C14A mutant. **j**, **k**, Representative western blots (**j**) and averaged data (**k**) showing that inhibition of proteasome by β -lac ($5 \mu\text{M}$ for 12 h) blocked the MG53-mediated suppression of phosphorylation of insulin receptor- β and IRS1 and Akt, and downregulation of insulin receptor- β and IRS1 in C2C12 myotubes. In **i** and **k**, phosphorylated and total proteins were normalized to GAPDH and presented as fold of respective control. For insulin stimulation, cells were pre-treated with insulin ($0.1 \mu\text{M}$) for 10 min ($n = 3$ for all groups). **l**, β -Lac abrogated MG53-mediated suppression of 2-NBDG uptake in C2C12 myotubes ($n = 8$). Data are mean \pm s.e.m. (** $P < 0.01$ as indicated).

contrast, the protein abundance of skeletal muscle IRS2, GLUT1 or GLUT4, which are also involved in glucose homeostasis, was unaffected by either MG53 overexpression or its ablation (Supplementary Fig. 17), indicating that they are not substrates for MG53.

Sequence analysis predicted that the cysteine-rich RING finger domain in the N terminus of MG53 and, in particular, the first cysteine residue (cysteine 14) which binds Zn^{2+} may be crucial for MG53 E3 ligase activity in light of previous investigations into other RING-type E3 ligases²⁹. Indeed, the RING finger deletion (Δ RING) or the alanine substitution of cysteine 14 (C14A) abolished or markedly attenuated the effects of MG53 on insulin receptor and IRS1 ubiquitination (Fig. 5f, g). Notably, these MG53 mutants affected neither the protein levels of the insulin receptor and IRS1 nor the insulin-induced activation of the insulin receptor-IRS1-Akt-GSK3- β signalling cascade (Fig. 5h, i and Supplementary Fig. 18). Thus, at the molecular level, the RING finger domain is required for MG53 E3 ligase activity. Furthermore, proteasome inhibition by clasto-lactacystin β -lactone (β -lac) abolished MG53-induced downregulation of the insulin receptor and IRS1, restored tyrosine phosphorylation of the insulin receptor and IRS1 (Fig. 5j, k) and Ser473 phosphorylation of Akt (Supplementary Fig. 19d) as well as insulin-induced glucose uptake (Fig. 5l), indicating the involvement of the proteasome system in MG53-mediated suppression of insulin signalling. Similar results were obtained with another proteasome inhibitor, MG132 (Supplementary Fig. 19a–c). Taken together, our *in vivo* and *in vitro* data demonstrate that MG53 acts as a novel E3 ligase that directly regulates the insulin receptor and IRS1 protein stability through ubiquitin-dependent degradation. This finding identifies MG53 as a mechanism underlying the simultaneous downregulation of the insulin receptor and IRS1 in the context of systemic insulin resistance and metabolic diseases.

In summary, whereas MG53 is known for its roles in membrane repair and cardioprotection, we have now shown that MG53 is universally upregulated in animal models with insulin resistance and metabolic disorders, and that muscle-specific MG53 upregulation is necessary and sufficient to trigger whole-body insulin resistance and metabolic syndrome. Mechanistically, MG53 acts as a novel muscle-specific E3 ligase targeting both the insulin receptor and IRS1 for ubiquitin-dependent degradation, hence constituting a crucial negative regulator of insulin signalling in skeletal muscle which, in turn, triggers systemic defects in insulin signalling and metabolism. These findings not only define MG53 as a powerful negative regulator of skeletal muscle insulin sensitivity, but also establish MG53-mediated suppression of muscle insulin signalling as a central mechanism underlying whole-body insulin resistance and metabolic syndrome, marking MG53 E3 ligase as a potentially important therapeutic target for the treatment of diverse metabolic diseases, including obesity, type 2 diabetes and associated cardiovascular complications.

METHODS SUMMARY

Blood analysis, extraction of mRNA and proteins, real-time PCR, western blot, co-immunoprecipitation, haematoxylin and eosin staining and immunofluorescence analysis were performed following standard protocols as previously described¹⁰. Blood pressure was measured with a BP-2000 Blood Pressure Analysis System (Visitech Inc.). Detailed information on *MG53^{-/-}* and *MG53^{Tg}* mice, HFD treatment, cell culture, antibodies, primers and probes are given in Methods.

Statistics. Differences between groups were examined for statistical significance using one-way analysis of variance (ANOVA) or repeated-measures ANOVA, when appropriate, with the Bonferroni post-test. All *P*-values below 0.05 were considered significant. Data are expressed as mean \pm s.e.m.

Full Methods and any associated references are available in the online version of the paper.

Received 19 October 2010; accepted 11 December 2012.

Published online 27 January 2013.

1. Eckel, R. H., Grundy, S. M. & Zimmet, P. Z. The metabolic syndrome. *Lancet* **365**, 1415–1428 (2005).

2. DeFronzo, R. A. *et al.* The effect of insulin on the disposal of intravenous glucose. Results from indirect calorimetry and hepatic and femoral venous catheterization. *Diabetes* **30**, 1000–1007 (1981).
3. Shulman, G. I. *et al.* Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by ^{13}C nuclear magnetic resonance spectroscopy. *N. Engl. J. Med.* **322**, 223–228 (1990).
4. McMillen, I. C. & Robinson, J. S. Developmental origins of the metabolic syndrome: prediction, plasticity, and programming. *Physiol. Rev.* **85**, 571–633 (2005).
5. Grundy, S. M. Metabolic syndrome: connecting and reconciling cardiovascular and diabetes worlds. *J. Am. Coll. Cardiol.* **47**, 1093–1100 (2006).
6. DeFronzo, R. A. & Tripathy, D. Skeletal muscle insulin resistance is the primary defect in type 2 diabetes. *Diabetes Care* **32** (Suppl. 2), S157–S163 (2009).
7. Lillioja, S. *et al.* Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. *N. Engl. J. Med.* **329**, 1988–1992 (1993).
8. Martin, B. C. *et al.* Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *Lancet* **340**, 925–929 (1992).
9. Cai, C. *et al.* MG53 nucleates assembly of cell membrane repair machinery. *Nature Cell Biol.* **11**, 56–64 (2009).
10. Cao, C. M. *et al.* MG53 constitutes a primary determinant of cardiac ischemic preconditioning. *Circulation* **121**, 2565–2574 (2010).
11. Fu, J. *et al.* Oleylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR- α . *Nature* **425**, 90–93 (2003).
12. Rao, R. H. Insulin resistance in spontaneously hypertensive rats. Difference in interpretation based on insulin infusion rate or on plasma insulin in glucose clamp studies. *Diabetes* **42**, 1364–1371 (1993).
13. Zhang, X. *et al.* Rhesus macaques develop metabolic syndrome with reversible vascular dysfunction responsive to pioglitazone. *Circulation* **124**, 77–86 (2011).
14. Coleman, D. L. Obese and diabetes: two mutant genes causing diabetes-obesity syndromes in mice. *Diabetologia* **14**, 141–148 (1978).
15. Saltiel, A. R. & Kahn, C. R. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799–806 (2001).
16. Thirone, A. C., Huang, C. & Klip, A. Tissue-specific roles of IRS proteins in insulin signaling and glucose transport. *Trends Endocrinol. Metab.* **17**, 72–78 (2006).
17. Kerouz, N. J., Horsch, D., Pons, S. & Kahn, C. R. Differential regulation of insulin receptor substrates-1 and -2 (IRS-1 and IRS-2) and phosphatidylinositol 3-kinase isoforms in liver and muscle of the obese diabetic (ob/ob) mouse. *J. Clin. Invest.* **100**, 3164–3172 (1997).
18. Goodyear, L. J. *et al.* Insulin receptor phosphorylation, insulin receptor substrate-1 phosphorylation, and phosphatidylinositol 3-kinase activity are decreased in intact skeletal muscle strips from obese subjects. *J. Clin. Invest.* **95**, 2195–2204 (1995).
19. Olefsky, J., Bacon, V. C. & Baur, S. Insulin receptors of skeletal muscle: specific insulin binding sites and demonstration of decreased numbers of sites in obese rats. *Metabolism* **25**, 179–191 (1976).
20. Capozza, F. *et al.* Caveolin-3 knockout mice show increased adiposity and whole body insulin resistance, with ligand-induced insulin receptor instability in skeletal muscle. *Am. J. Physiol. Cell Physiol.* **288**, C1317–C1331 (2005).
21. Zisman, A. *et al.* Targeted disruption of the glucose transporter 4 selectively in muscle causes insulin resistance and glucose intolerance. *Nature Med.* **6**, 924–928 (2000).
22. Brüning, J. C. *et al.* A muscle-specific insulin receptor knockout exhibits features of the metabolic syndrome of NIDDM without altering glucose tolerance. *Mol. Cell* **2**, 559–569 (1998).
23. Laustsen, P. G. *et al.* Essential role of insulin and insulin-like growth factor 1 receptor signaling in cardiac development and function. *Mol. Cell. Biol.* **27**, 1649–1664 (2007).
24. Nakao, R. *et al.* Ubiquitin ligase Cbl-b is a negative regulator for insulin-like growth factor 1 signaling during muscle atrophy caused by unloading. *Mol. Cell. Biol.* **29**, 4798–4811 (2009).
25. Ramos, F. J., Langlais, P. R., Hu, D., Dong, L. Q. & Liu, F. Grb10 mediates insulin-stimulated degradation of the insulin receptor: a mechanism of negative regulation. *Am. J. Physiol. Endocrinol. Metab.* **290**, E1262–E1266 (2006).
26. Rui, L., Yuan, M., Frantz, D., Shoelson, S. & White, M. F. SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2. *J. Biol. Chem.* **277**, 42394–42398 (2002).
27. Xu, X. *et al.* The CUL7 E3 ubiquitin ligase targets insulin receptor substrate 1 for ubiquitin-dependent degradation. *Mol. Cell* **30**, 403–414 (2008).
28. Shi, J., Luo, L., Eash, J., Ibebunjo, C. & Glass, D. J. The SCF-Fbxo40 complex induces IRS1 ubiquitination in skeletal muscle, limiting IGF1 signaling. *Dev. Cell* **21**, 835–847 (2011).
29. Joazeiro, C. A. *et al.* The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. *Science* **286**, 309–312 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank H. P. Cheng, G. Feng, X. Fu and L. P. Wei for discussions, and S. L. Guo, T. Zhang, X. H. Wang, D. Y. Chen, J. Y. Peng, L. Huang, W. Q. Zhang, N. Hou, L. Pan, L. Chen and Y. L. Liu for their technical support. Special thanks to H. Takeshima and J. J. Ma for their support in providing MG53^{-/-} mice. This work was supported by the National Basic Research Program of China (2012CB518000, 2013CB531200, 2012CB944501) and the National Natural Science Foundation of China (81070674, 81070116, 3/22/002 and 81130073).

Author Contributions R.S., W.P. and Y.Z. are equally contributing first authors. R.S. generated the initial idea and conducted key experiments. R.S., W.P., Y.Z., C.-M.C. and R.-P.X. designed the study, analysed the data and wrote the manuscript. C.-M.C. and R.-P.X. interpreted significance of the study. R.S., W.P., Y.Z., F. Lv, H.-K.W., J.G., Y.C., Y.P., Xin Z., L.J., M.Z., Pe.J., F. Liu and S.M. performed the experiments. Pi.J. helped in the generation of MG53 transgenic mice. Xiu.Z. provided the nonhuman primate tissues.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.-P.X. (xiaor@pku.edu.cn) or C.-M.C. (caochunmei@pku.edu.cn).

METHODS

Reagents and materials. Antibodies to PY100, p-Akt, Akt, p-GSK3- β and IRS2 were from Cell Signaling Technology; anti-MG53 and GLUT1 were from Abcam; p-IR- β and IRS1 antibodies were from Upstate; IR- β , haemagglutinin (HA), GAPDH, GSK3- β and GLUT4 antibodies were from Santa Cruz Biotechnology; MG132, clasto-lactacystin β -lactone (β -lac), anti- β -actin, Flag, Myc and insulin antibodies were from Sigma-Aldrich. 2-(N-(7-nitrobenz-2-oxa-1,3-diazol-4-yl) amino)-2-deoxyglucose (2-NBDG) was from Invitrogen. Human IGF-I was from Peptrotech. Unless indicated otherwise, all chemicals were from Sigma-Aldrich.

Animal models with insulin resistance. *db/db* mice (male *db/db* mice at 25 weeks of age) and lean control mice were from the Jackson Laboratory. Spontaneously hypertensive rats (male SHR at 12 months of age) and Wistar Kyoto rats (WKY) were from Vital River Laboratories, Beijing, China. The development and characterization of a nonhuman primate (NHP) model of spontaneous insulin resistance and metabolic syndrome was reported previously¹³.

MG53^{-/-} mice and dietary intervention. All animal procedures and euthanasia were performed in accordance with protocols approved by the Committee for Animal Research of Peking University, China, and conformed to the Guide for the Care and Use of Laboratory Animals (NIH publication No. 86-23, revised 1985). All mice were maintained in a temperature-controlled barrier facility with a 12-h light/dark cycle and were given free access to food and water in the Center for Experimental Animals at Peking University, Beijing, China (an AAALAC-accredited experimental animal facility). Only male animals were used in this study. The generation of MG53^{-/-} mice was described previously¹⁰. Dietary intervention with a high-fat diet (60% calories from fat, Cat. # D12492, Research Diets Inc.)¹¹ or a chow diet (11.4% calories from fat, Academy of Military Medical Sciences, China) started from 3 weeks of age in MG53^{-/-} mice and wild-type littermates.

MG53^{tg} mice. The full-length murine MG53 cDNA coding sequence was cloned into the XhoI site of pUC-CAGGS, under the control of the chicken β -actin promoter. After linearization with SalI and subsequent gel-purification, this construct was microinjected into the pronuclei of fertilized C57BL/6J mouse eggs. PCR was used for genotyping.

Plasmid and adenoviral vectors. DNA fragments corresponding to full-length or Δ RING (deletion of the RING domain) MG53 were amplified from a mouse cDNA library by PCR and inserted into p3 \times FLAG-CMV-10 Expression Vector (Sigma-Aldrich) using the BglII and XbaI restriction sites. Full-length MG53 sequence was also inserted into pcDNA4/TO/Myc-His B expression vector (Invitrogen) between the KpnI and XhoI restriction sites. The C14A MG53 mutant (the 14th cysteine substituted by alanine) was generated from the wild-type MG53 construct (FL-MG53) by point mutation using Stratagene's QuikChange II site-directed mutagenesis kit. The insulin receptor sequence was subcloned from pBABE-bleo human insulin receptor B (Addgene) and inserted into pcDNA4/TO/Myc-His B expression vector (Invitrogen) between the HindIII and XbaI restriction sites. IRS1 was subcloned from pBS mouse IRS1 (Addgene) and inserted into pcDNA4/TO/myc-His B expression vector (Invitrogen) between the HindIII and NotI restriction sites. The constructs expressing N-terminal HA-tagged ubiquitin and C-terminal Flag-tagged insulin receptor were provided by D. Chen and I. Leibiger, respectively. Adenovirus expressing GFP or GFP-MG53 was described previously¹⁰.

Cell culture, adenoviral infection and plasmid transfection. C2C12 myoblasts (from Cell Resource Center, IBMS, CAMS/PUMC) were cultured at 37 °C under 5% CO₂ in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Sigma-Aldrich), 0.11 g l⁻¹ sodium pyruvate, and 1% penicillin-streptomycin. When C2C12 myoblasts reached 90% confluence, we performed gene transfer by adenoviral infection or plasmid transfection. After gene transfer, cells were cultured in DMEM (2% horse serum) for 4 days to differentiate into myotubes.

Co-immunoprecipitation, western blot and confocal immunofluorescent imaging. Tissues or cells were lysed in lysis buffer A (30 mM HEPES at pH 7.6, 100 mM NaCl, 0.5% Nonidet P-40, and protease inhibitors mixture) on ice for 10 min, and the lysates were centrifuged at 13,000 r.p.m. for 10 min. The co-immunoprecipitation and western blot assays were performed as described previously¹⁰. Analysis of tyrosine phosphorylation of IRS1 was performed by immunoprecipitation of IRS1 with anti-IRS1 from total lysate, followed by western blot with anti-pTyr antibody (PY100). Other proteins were analysed with specific antibodies. Immunofluorescent staining was performed with histological sections of pancreas using anti-insulin antibody, and images were made with a laser scanning confocal microscope (LSM-700, Zeiss), as previously described¹⁰.

Ubiquitination assay. C2C12 myotubes transfected with the indicated plasmids were treated with 10 μ M MG132 for 12 h before collection, then the cells were rinsed in ice-cold PBS, and lysed with RIPA buffer (in mM: 200 NaCl, 20 Tris-Cl at pH 8.0, 1 EDTA, 1 EGTA, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS, 2.5 sodium pyrophosphate, 1 β -glycerol phosphate, 1 Na₃VO₄, protease

inhibitor mixture and 10 μ M MG132) on ice for 10 min, followed by centrifugation at 13,000 r.p.m. for 10 min. The supernatant was pre-cleaned with nProtein A Sepharose 4 Fast Flow (GE Healthcare) for 1 h, and incubated with anti-Myc antibody and protein A agarose beads for 4 h at 4 °C. The resins were then washed three times with RIPA buffer and resolved onto SDS-PAGE for western blot.

To assay *in vivo* ubiquitination, skeletal muscles were ground into powders in liquid nitrogen, lysed with RIPA buffer mentioned above for 1 h at 4 °C, and then centrifuged at 13,000 r.p.m. for 10 min. Protein A agarose beads were incubated with indicated antibody for 8 h at 4 °C, and followed by incubation with a total of 700 μ g lysate for another 3 h at 4 °C. The immunoprecipitated beads were extensively washed in RIPA buffer, eluted with 2 \times SDS-sample buffer and analysed by western blots.

Gene expression analysis and primers. The following primer pairs were used for quantitative real-time PCR: 18S RNA, 5'-GGAAGGGCACCAGGAGT-3' (forward) and 5'-TGCAGCCCCGACATCTAAG-3' (reverse). The primers for rat MG53 were 5'-CGAGCAGGACCGCACACTT-3' (forward) and 5'-CCAGGAACATCCGCATCTT-3' (reverse). The primers for mouse MG53 were 5'-TGTGTGCCTCGCTCGGTTC-3' (forward) and 5'-TCTGCTTCACGGTCC AGAGAA-3' (reverse). The primers for mouse insulin receptor were 5'-GGCAG TTTGTGGAACGGTGCT-3' (forward) and 5'-GACCATCCAGATAGAAGT TGCGACA-3' (reverse). The primers for rat insulin receptor were 5'-GTCTT CACTGCTTCTCCGACAT-3' (forward) and 5'-AGCCTTGTCTCCTCGCT GTA G-3' (reverse). The primers for mouse *Irs1* were 5'-GCAGCAATGAG GGCAACTC-3' (forward) and 5'-TGCTTGGCACAAATGTAGAACGTG-3' (reverse). The primers for rat *IRS1* were 5'-GAAGAAGTGGCGGCACAAGT-3' (forward) and 5'-GTCAGGCAGAGGCGGTAGAT-3' (reverse). Amplification was performed as follows: 94 °C for 30 s and 30 cycles at 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s. The cycle number at which the emission intensity of the sample rose above baseline was referred to as the Ct value (threshold cycle) and was proportional to target concentration. Data presented are the average of at least four independent experiments.

Northern blot. Total RNA was isolated, fractionated by electrophoresis on 1.2% agarose/6% formamide gels, and transferred to Hybond nylon membranes (Amersham Biosciences). Probe labelling, hybridization, stripping and reprobing were carried out as previously described³⁰. A 620-base-pair cDNA template of mouse MG53 used for the probe was generated by PCR using the primers 5'-CTGTGCTTGACGCTGTTCGAT-3' and 5'-ATCTGCCTCAGTTGCTCTAGG TAAG-3'.

Histological analysis. The tissues (liver, pancreas, visceral fat, brown fat and skeletal muscle) for histological analysis were fixed in 4% paraformaldehyde (pH 7.4) overnight, embedded in paraffin, and serially sectioned at 5 μ m. Standard haematoxylin and eosin staining or immunofluorescent staining was performed on these sections.

Blood pressure measurement. Systolic and diastolic blood pressure was measured non-invasively in conscious mice, using a tail-cuff system warmed to 37 °C (Visitech BP-2000 Blood Pressure Analysis System). Mice were habituated to the device for 7–10 days and underwent two cycles of 10 measurements per day for 3 days for blood pressure determination.

2-NBDG uptake assay. C2C12 myotubes with indicated gene transfer were incubated with serum-free DMEM for 12 h, and then maintained in Krebs-Ringer phosphate buffer (in mM: 128 NaCl, 1.4 CaCl₂, 1.4 MgSO₄, 5.2 KCl, 10 Na₂HPO₄, and 2 sodium pyruvate, pH 7.4) for 30 min at 37 °C, and subsequently treated with 0.1 μ M insulin and 100 μ M 2-NBDG (Invitrogen) for 6 h at 37 °C. Then, the cells were washed three times in ice-cold PBS, digested with 0.5% trypsin, centrifuged at 1,000 r.p.m. for 5 min, and then re-suspended in PBS. Finally, the FL1 fluorescence value was measured in a FACSCalibur Flow Cytometer (BD)³¹.

Metabolic measurements. For GTTs, mice were fasted overnight (for 16 h) and then injected intraperitoneally (i.p.) with D-glucose (2 g kg⁻¹ body weight). For ITTs, mice were randomly fed and injected i.p. with bovine insulin (0.75 U kg⁻¹ body weight, Sigma-Aldrich). To assay glucose-stimulated insulin release, mice fasted for 16 h were injected i.p. with D-glucose (2 g kg⁻¹). We collected blood from a tail vein before injection and at different time points after injection (as indicated in the figures). Glucose and insulin concentrations were measured with an AccuCheck blood glucose meter (Roche Diagnostics Inc.) and ELISA kits from Linco Research (catalogue number EZRMI-13K), respectively. Serum triglyceride and cholesterol concentrations were measured with kits from Wako Diagnostics (catalogue number 290-63701 and 294-65801, respectively).

For metabolic rate analysis, mice were housed individually under a 12-h light/dark cycle. A comprehensive animal metabolic monitoring system (CLAMS; Columbus Instruments) was used to evaluate oxygen consumption (VO₂) and carbon dioxide production (VCO₂) continuously over a 72-h period. Energy expenditure was calculated using the formula: energy expenditure = (3.815 + 1.232 VO₂/VCO₂) \times VO₂.

Statistics. Differences between groups were examined for statistical significance using one-way analysis of variance (ANOVA) or repeated-measures ANOVA, when appropriate, with the Bonferroni post-test. All *P*-values below 0.05 were considered significant. Data are expressed as mean \pm s.e.m.

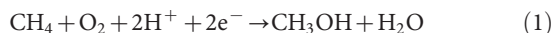
30. Huang, H. *et al.* Profiling of mismatch discrimination in RNAi enabled rational design of allele-specific siRNAs. *Nucleic Acids Res.* **37**, 7560–7569 (2009).
31. Zhong, L. *et al.* The histone deacetylase Sirt6 regulates glucose homeostasis via Hif1 α . *Cell* **140**, 280–293 (2010).

Control of substrate access to the active site in methane monooxygenase

Seung Jae Lee¹, Michael S. McCormick¹, Stephen J. Lippard¹ & Uhn-Soo Cho²

Methanotrophs consume methane as their major carbon source and have an essential role in the global carbon cycle by limiting escape of this greenhouse gas to the atmosphere^{1–3}. These bacteria oxidize methane to methanol by soluble and particulate methane monooxygenases (MMOs)^{1–4}. Soluble MMO contains three protein components, a 251-kilodalton hydroxylase (MMOH), a 38.6-kilodalton reductase (MMOR), and a 15.9-kilodalton regulatory protein (MMOB), required to couple electron consumption with substrate hydroxylation at the catalytic diiron centre of MMOH². Until now, the role of MMOB has remained ambiguous owing to a lack of atomic-level information about the MMOH–MMOB (hereafter termed H–B) complex. Here we remedy this deficiency by providing a crystal structure of H–B, which reveals the manner by which MMOB controls the conformation of residues in MMOH crucial for substrate access to the active site. MMOB docks at the $\alpha_2\beta_2$ interface of $\alpha_2\beta_2\gamma_2$ MMOH, and triggers simultaneous conformational changes in the α -subunit that modulate oxygen and methane access as well as proton delivery to the diiron centre. Without such careful control by MMOB of these substrate routes to the diiron active site, the enzyme operates as an NADH oxidase rather than a monooxygenase⁵. Biological catalysis involving small substrates is often accomplished in nature by large proteins and protein complexes. The structure presented in this work provides an elegant example of this principle.

Bacterial multicomponent monooxygenases (BMMs) are unique among diiron proteins by virtue of their ability to hydroxylate a broad spectrum of hydrocarbon substrates^{1–3}. Soluble MMOs, alkene monooxygenases, phenol hydroxylases and alkene/aromatic monooxygenases belong to the BMM superfamily^{1,2,4}. Soluble MMO is the only BMM capable of catalysing the conversion of methane selectively to methanol by activation of molecular oxygen for insertion of an oxygen atom into a C–H bond (104.9 kcal mol^{–1}), as illustrated in equation (1)^{2,3}:



The crystal structure of MMOH revealed a dimeric ($\alpha_2\beta_2\gamma_2$) architecture with a glutamate-bridged diiron active site in each α -subunit^{6,7}. Substrate turnover in soluble MMO is initiated by electron transfer from MMOR to the resting state diiron(III) hydroxylase MMOH_{ox}, converting it to the reduced diiron(II) state, MMOH_{red}. In the presence of MMOB, oxygen activation at the active site of MMOH_{red} yields a diiron(III) peroxo intermediate that rapidly decays to form Q, the diiron(IV) species that performs methane oxidation, returning the enzyme to the resting state^{8,9}. Comparison of oxidized and reduced hydroxylase structures revealed charge neutral active sites, reduction being accompanied by conversion of two bridging OH[–] ions to water^{6,7,10} (Supplementary Fig. 1). When MMOB binds MMOH, the Fe(III)₂ → Fe(II)₂ reduction potential is lowered, but only in the absence of MMOR¹¹. Binding of MMOB to MMOH increases the rate and specificity of substrate hydroxylation^{12–14}. The nature of the internal MMOH conformational changes induced by MMOB has

remained unknown owing to the absence of structural information about the complex formed between these two component proteins.

A crystal of H–B that diffracted to 2.9 Å resolution allowed us to determine the X-ray structure by molecular replacement, as outlined in the Methods and Supplementary Table 1. There are two H–B complexes in the asymmetric unit comprising four crystallographically independent $\alpha\beta\gamma$ protomers (Supplementary Fig. 2). Within individual dimers, the protomers are related by a non-crystallographic two-fold symmetry axis (Fig. 1a and Supplementary Fig. 3) and have nearly identical overall structures (Supplementary Table 2). MMOB binds to the hydroxylase with its core residues (Asp 36–Leu 129) located primarily in a ‘canyon’ region⁷ formed at the $\alpha_2\beta_2$ interface of the two MMOH protomers. Similar canyon motifs occur in the hydroxylase components of phenol hydroxylase and toluene monooxygenases for binding their respective regulatory proteins^{15–17}, but these proteins lack the amino-terminal tail that is critical for the function of soluble MMO (Supplementary Fig. 4). Proof that MMOB binds in the canyon of MMOH—the archetypal and most investigated member of the BMM family—and the structure and function of the MMOB amino terminus are provided for the first time, to our knowledge, by the present structure determination.

NMR spectroscopic analysis¹⁸ of unbound MMOB from *Methylococcus capsulatus* (Bath) revealed a compact core region (Fig. 1b) and an unstructured N-terminal tail that is ~35 amino acids longer than the corresponding region of regulatory proteins from all other BMM subclasses (Supplementary Fig. 4). In the H–B complex, the MMOB core exhibits only minor structural changes with respect to that in the unbound protein, as reflected by C α root mean squared deviation values of ~2.1 Å. The N terminus of MMOB becomes very well ordered in H–B, forming a remarkable ring-shaped structure on the α -subunit of MMOH (Fig. 1a, b). The extended N terminus in MMOB was previously noted to be crucial for soluble MMO catalysis^{19,20}, and those results are confirmed in the present study, in which N-terminal truncates (Δ 1–8, Δ 1–17 and Δ 1–33) showed substantially reduced activity with respect to full-length MMOB (Fig. 1c).

The MMOB N terminus binds to helices H and 4 of MMOH in the complex through hydrogen-bonding as well as hydrophobic interactions (Supplementary Fig. 5). A small α -helix (Gly 17–Phe 25) in the MMOB tail facilitates formation of its ring-shaped structure on the MMOH surface (Fig. 1a and Supplementary Fig. 5). Within this ring structure, Phe residues 20, 24 and 25 of MMOB generate hydrophobic interactions with Lys 303 (helix H), Val 302 and Tyr 340 (helix 4 of MMOH), respectively, as shown in Supplementary Fig. 5. These features of the N terminus may help to anchor MMOB on the MMOH surface, making it difficult for MMOR to displace it from a preformed H–B complex. Such a consequence would account for the diminished rate of intermolecular electron transfer observed between MMOR and MMOH in preformed H–B²¹. In addition, hydrophilic residues including Lys 18, Asp 19, Asp 22 and Gln 23 of MMOB, located opposite the H–B binding interface, contribute to the solubility of the H–B complex.

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.

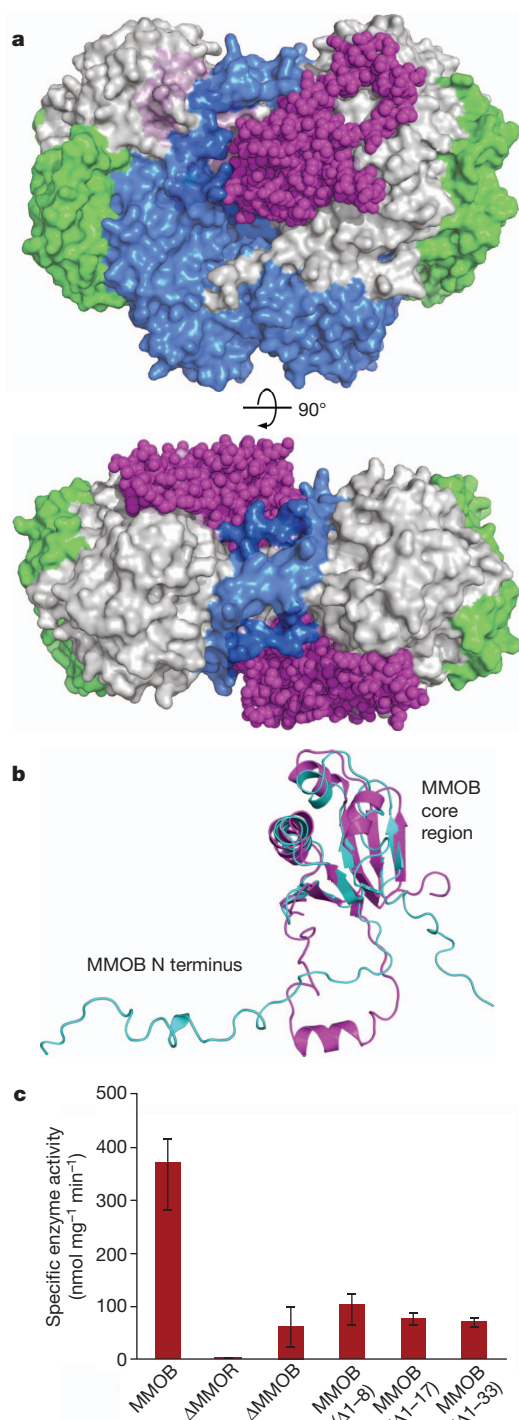


Figure 1 | MMOB induces conformational changes that affect function.

a, Front (top) and top (bottom) views of a space-filling representation of the H-B complex. MMOB (magenta) binds to the canyon region formed by the α - (grey) and β - (blue) subunits of MMOH. The MMOH γ -subunit is depicted in green. **b**, Structural alignment of the solution NMR structure of MMOB from *M. capsulatus* (Bath) (PDB accession 1CKV) (cyan) with that of MMOB in the H-B complex (magenta). An α -helix (Gly 17–Phe 25) forms in the MMOB N terminus after complex formation with MMOH. **c**, Soluble MMO activity assay with wild-type and truncated versions of MMOB. Propylene is converted to propylene oxide in the presence of NADH. Native MMOB is required for maximum soluble MMO activity. N-terminal truncated MMOB constructs tested, (Δ 1–8), (Δ 1–17) and (Δ 1–33), show similar activity profiles to that observed in the absence of MMOB ($n = 3$, mean \pm s.d.).

Furthermore, when MMOB docks onto the α -subunit of MMOH, it imparts important conformational changes in the hydroxylase. These structural changes are largely confined to the α -subunits and involve particularly helices E, F, H and 4 (Supplementary Figs 6 and 7). In the H-B complex, Tyr 8 and Ser 111 of MMOB allosterically induce considerable amino acid side-chain movements near the diiron active site in MMOH helix E. Tyr 8 forms hydrogen bonds with Arg 307 and Glu 299 in MMOH helix H reorienting Trp 308 (Fig. 2a, b). This reorientation of Trp 308 is stabilized by π -interactions with Tyr 76 and Trp 78 (β 3-strand) of MMOB. In addition, Ser 111 of MMOB forms a hydrogen bond with Asn 214 in MMOH helix E, which triggers a side-chain reorientation in Thr 213, an active site, second coordination sphere residue of importance for the formation of oxygenated intermediates in the catalytic cycle^{23,9} and possibly proton-coupled electron transfer. In H-B, the conformational change of Thr 213 generates hydrogen bonds with Glu 240. This event closes a pore in the MMOH structure, the shortest access route between the diiron active site and the protein surface defined by residues Glu 240, Thr 213 and Asn 214 (Figs 2c, d and 3). This pore was previously proposed to be involved in proton transfer²².

Protons are an important substrate in BMM catalytic cycles^{9,22–24}, and the H-B structure provides insight into the role that MMOB may have in facilitating proton access to the catalytic diiron centre in soluble MMO. In the H-B complex, the conformational change of Thr 213 is accompanied by formation of a bifurcated hydrogen bond between the hydroxyl group of this residue and the carboxylate side chain of Glu 240. Glu 243 simultaneously undergoes a 'carboxylate shift'²⁵ (vide infra). In the absence of MMOB, the Glu 240 and Asn 214 side chains in MMOH are solvent accessible and linked through hydrogen bonding to a water molecule or hydronium ion (Fig. 2c). After H-B complex formation (Fig. 2d), Glu 240 shifts towards the protein interior possibly delivering a proton in the process. One possible model is that, during the O₂ activation steps by MMOH to form the peroxo and Q intermediates, both of which require a proton transfer^{9,23}, MMOB core binding, release and rebinding to the hydroxylase might facilitate delivery of the requisite two solvent-derived protons through the pore. The anchoring of MMOB by its N terminus may allow the core to function in this manner without complete dissociation of the regulatory protein from the hydroxylase. The presence of these protons in the active site would also facilitate product and hydroxide ion release during reduction of the diiron(III) centre in MMOH_{ox} to form MMOH_{red}. Delivering protons through the pore may be one of the primary functions of the regulatory proteins in the BMM family.

Yet another important feature of MMOB binding to the α -subunit of MMOH is to control methane and O₂ access to the active site. Previous structural analyses of MMOH crystals soaked in solutions of halogenated substrate analogues or pressurized with Xe identified a putative access route for these substrates^{26,27}. In the MMOH_{ox} structure, cavities 2 and 3 are connected, but there is a discontinuity between cavities 1 and 2. This break in freely diffusible space blocks access of methane and oxygen to the active site (Fig. 3a and Supplementary Fig. 8). Molecular access to the diiron site via the cavities is gated by residues Phe 188 and Leu 110 (refs 7, 22, 26). In the H-B complex, cavities 1 and 2 become connected as a consequence of a change in conformation of the Phe 188 side chain, and a structural comparison of MMOH_{ox} with that of H-B highlights the difference (Fig. 3 and Supplementary Figs 8 and 9). A major function of MMOB binding to MMOH is, therefore, to facilitate methane and O₂ access to the diiron active site by opening the gate. It is noteworthy that this structural alteration occurs concomitantly with closure of the pore (Fig. 3 and Supplementary Figs 8 and 9). Opening of the pore upon MMOB dissociation also supports its previously proposed role as a hydrophilic route for methanol release²⁶.

Changes also occur in the geometry of the diiron centre after MMOB binding, in accord with spectroscopic studies that revealed

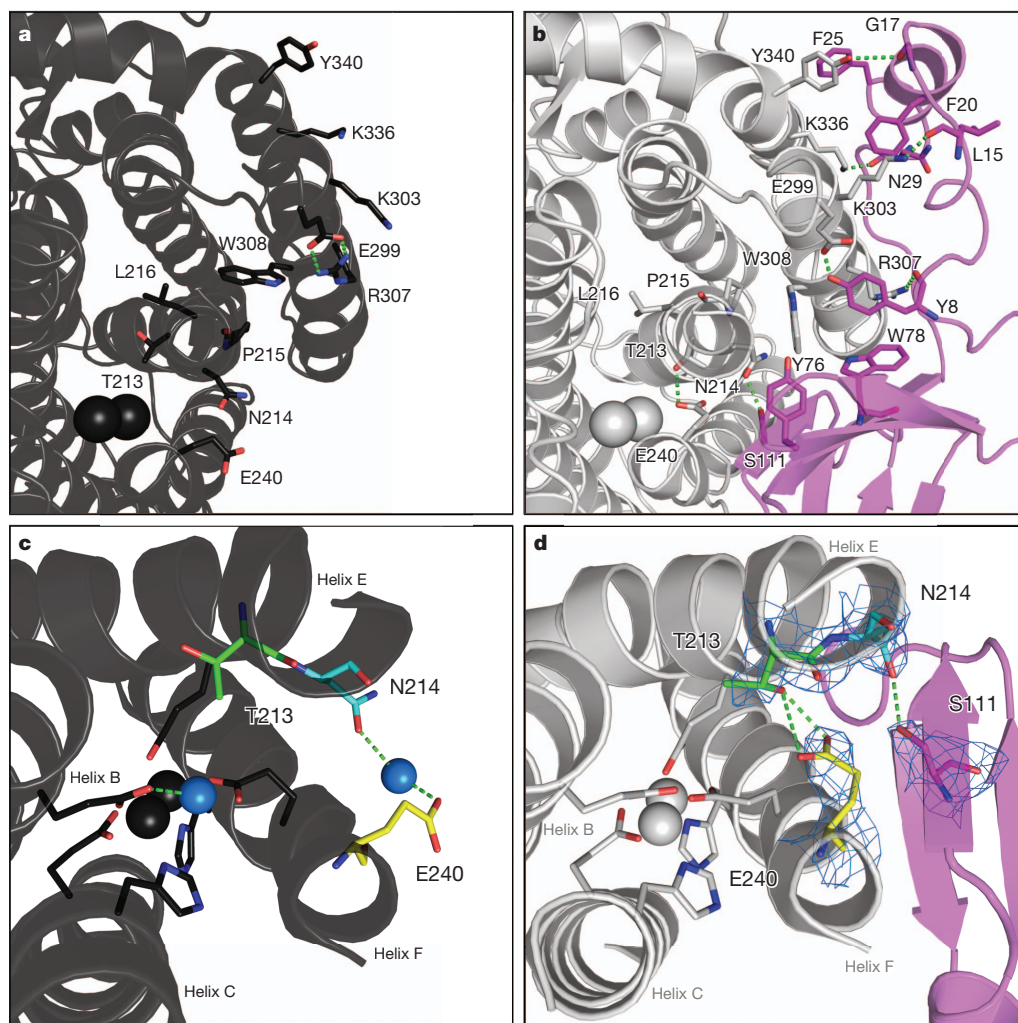
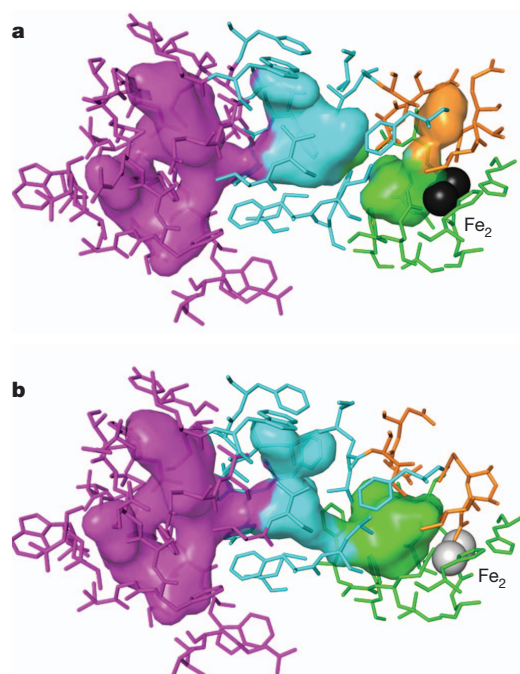


Figure 2 | Conformational changes near the diiron centre and pore residues in MMOH after MMOB binding. **a, b,** Protein backbones are shown as ribbons in black (MMOH_{ox}), grey (MMOH from the H–B complex) and magenta (MMOB from the H–B complex). Iron atoms are depicted as black (MMOH_{ox}) or grey (H–B) van der Waals spheres. Interactions between key residues at the protein–protein interface are depicted as sticks; carbon atoms are coloured to match the protein backbone from which they stem, nitrogen atoms are shown in blue, and oxygen atoms in red; hydrogen bonds are represented as green dashes. **c, d,** Conformational changes in the MMOH pore residues after MMOB binding are illustrated in MMOH_{ox} (**c**) and H–B (**d**). Iron-ligating helices B, C, E and F and iron atoms are shown as black (**c**) and grey (**d**) ribbons, and MMOB as magenta ribbons. Active site side-chain ligands are shown as sticks in grey. Residues Thr 213 (green), Asn 214 (cyan) and Glu 240 (yellow) in MMOH, and Ser 111 in MMOB (magenta), are rendered as sticks. Nitrogen and oxygen atoms are shown in blue and red, respectively. The blue spheres in **c** are water molecules or hydronium ion. $2F_o - F_c$ electron density at 1.0σ from the H–B structure is drawn as a light blue mesh about key residues. The Glu 240 carboxylate side chain may function to deliver protons from solvent to the diiron site while closing the pore (see text).



conformational rearrangements of coordinated amino acid side chains in the H–B complex^{2,3,9,14}. The coordination environments of the iron atoms in H–B (Fig. 4a and Supplementary Fig. 10) exhibit many similarities to, as well as some key differences from, those observed in MMOH_{ox} and MMOH_{red} (Fig. 4b, c and Supplementary Fig. 10)^{7,10,28}. As in the other structures of MMOH, Fe1 and Fe2 in H–B are positioned within the four-helix bundle formed by helices B, C, E and F. The positions of helices E and F shift after MMOB binding, moving Fe2 ~ 1.1 Å from its location in MMOH_{ox} (Supplementary Fig. 11). The coordination of Glu 243 resembles that in H_{red}, but the Fe \cdots Fe distance in H–B is closer to that in MMOH_{ox}.

Figure 3 | Pore closure and cavity opening after MMOH–MMOB complex formation. **a, b,** Views of cavities 1 (green), 2 (cyan) and 3 (magenta), and the pore region (orange) are shown as translucent van der Waals surfaces in the hydroxylase interior. Protein residues that contribute to the interior surfaces are shown as sticks and coloured to match the surfaces to which they primarily contribute. Active site iron atoms are depicted as black or grey spheres. MMOH_{ox} (PDB accession 1MTY) is shown in **a**, and the H–B complex is depicted in **b**. The pore region in the hydroxylase interior becomes completely occluded by structural rearrangements that occur in the iron-ligating helices E and F, and in residues Asn 214 and Glu 240 in particular, after complex formation with the regulatory protein. Cavity-gating residue Phe 188 also changes orientation as a consequence of the regulatory protein binding-induced rearrangements in helices E and F, and as a result cavities 1 and 2 (cyan/green interface) merge. Hydrophobic cavities 2 and 3 in MMOH have sufficient space to accommodate the binding and translocation of these two gaseous substrates.

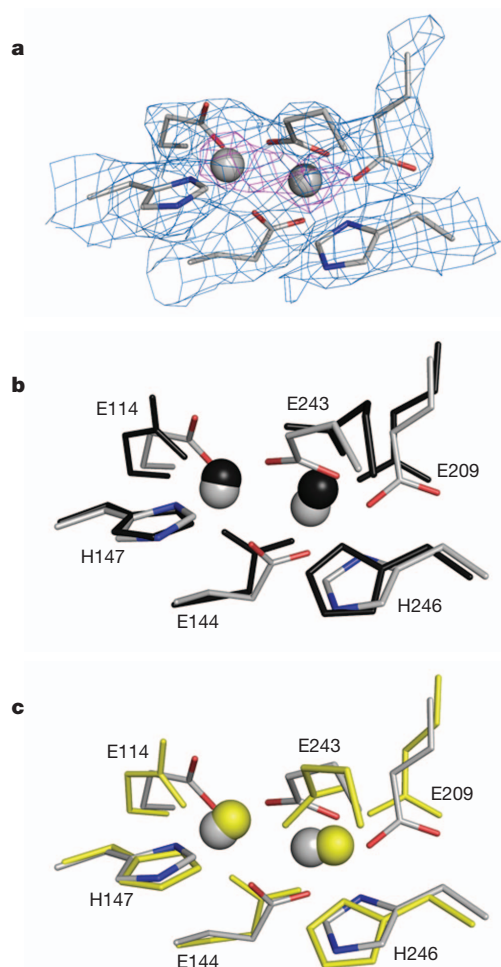


Figure 4 | Coordination geometry at the diiron active site of MMOH.

a, Electron density at the diiron site in H-B. Side-chain ligands are shown as sticks in grey (carbon), blue (nitrogen) and red (oxygen). $2F_o - F_c$ electron density at 1.0σ and 5.0σ is drawn as a mesh in light blue and magenta, respectively. **b**, **c**, Views comparing the diiron site in H-B (grey) with that in MMOH_{ox} (**b**, PDB accession 1MTY; black) and in MMOH_{red} (**c**, PDB accession 1FYZ; yellow) are also presented. After MMOB binding, Glu 243 undergoes a substantial conformational change, involving simultaneous chelation of Fe2 and bridging to Fe1. Such bidentate coordination of the Glu 243 side chain resembles that in MMOH_{red}, but distances between the carboxylate oxygen atoms and Fe2 (OE1-Fe2 and OE2-Fe2) are shorter in H-B (1.9 and 2.0 Å) than in MMOH_{red} (2.4 and 2.4 Å). This result is more consistent with an Fe(III) than an Fe(II) oxidation state for the iron atoms in H-B. Solvent-derived ligands, such as hydroxide ion and water, are not observed in the H-B complex, possibly owing to the 2.9 Å resolution.

Individual refinement of crystallographically independent diiron active sites within the four protomers revealed the same coordinated ligands, although with slightly different geometries (Supplementary Fig. 12). The Fe1 and Fe2 ions bond to the δ -N atoms of His 147 and His 246, respectively, Glu 144 bridges the two metals, and Glu 209 binds in a monodentate fashion to Fe2, all as in MMOH_{ox} structures (Fig. 4 and Supplementary Figs 10 and 12). The most notable change occurs in the Glu 243 side-chain carboxylate, which chelates Fe2 in a bidentate manner while being singly bonded to Fe1. In MMOH_{ox}, the carboxylate of Glu 243 forms a single bond with Fe2, and the dangling oxygen atom hydrogen bonds to a terminal water coordinated to Fe1 and a hydroxide ion bridging Fe1 and Fe2. The MMOH_{ox} active site contains a water molecule terminally bound to Fe1 that is not observed in H-B, either because it cannot be distinguished at 2.9-Å resolution or because it is not present in the H-B structure.

The present structure reveals how the MMOB regulatory protein controls substrate access to the diiron centre in the soluble MMO hydroxylase. Docking of the MMOB core in the MMOH canyon is accompanied by ordering of its long N-terminal tail on the α -subunit of the hydroxylase, triggering allosteric changes that control proton, methane and oxygen access to the active site. The timed entry of these substrates is important to assure events required for MMOR conversion of MMOH_{ox} to MMOH_{red} and the generation of oxygenated intermediates that react with methane during the oxygen activation and substrate hydroxylation steps of the catalytic cycle. In this manner, MMOB can function to couple the consumption of electrons with efficient hydrocarbon hydroxylation. Finally, the present results can be used as a leading example of the use by nature of a large protein complex to delineate access pathways of some of its smallest substrates to the active site of a metalloenzyme to achieve a remarkable catalytic reaction.

METHODS SUMMARY

M. capsulatus (Bath) cultures were grown by fermentation and MMOH was purified as described previously⁹. Recombinant full-length and truncated MMOB proteins were expressed and purified from *Escherichia coli* as described^{29,30}. Crystallization, crystal structure determination and enzyme activity studies with the MMOB N-terminal deletion mutants were performed as described in Methods. Data collection was performed at the Advanced Light Source beamline 8.2.2. at the Lawrence Berkeley National Laboratory, and the structure was determined by molecular replacement using the program Phaser with MMOH_{ox} (Protein Data Bank (PDB) accession 1MTY) and MMOB (PDB accession 1CKV) as search models.

Full Methods and any associated references are available in the online version of the paper.

Received 24 July; accepted 21 December 2012.

Published online 10 February 2013.

- Hanson, R. S. & Hanson, T. E. Methanotrophic bacteria. *Microbiol. Rev.* **60**, 439–471 (1996).
- Merkx, M. *et al.* Dioxygen activation and methane hydroxylation by soluble methane monooxygenase: A tale of two irons and three proteins. *Angew. Chem. Int. Ed. Engl.* **40**, 2782–2807 (2001).
- Waller, B. J. & Lipscomb, J. D. Dioxygen activation by enzymes containing binuclear non-heme iron clusters. *Chem. Rev.* **96**, 2625–2658 (1996).
- Leahy, J. G., Batchelor, P. J. & Morcomb, S. M. Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* **27**, 449–479 (2003).
- Gassner, G. T. & Lippard, S. J. Component interactions in the soluble methane monooxygenase system from *Methylococcus capsulatus* (Bath). *Biochemistry* **38**, 12768–12785 (1999).
- Elango, N. *et al.* Crystal structure of the hydroxylase component of methane monooxygenase from *Methylosinus trichosporium* OB3b. *Protein Sci.* **6**, 556–568 (1997).
- Rosenzweig, A. C., Frederick, C. A., Lippard, S. J. & Nordlund, P. Crystal structure of a bacterial non-haem iron hydroxylase that catalyzes the biological oxidation of methane. *Nature* **366**, 537–543 (1993).
- Shu, L. *et al.* An Fe₂^{IV}O₂ diamond core structure for the key intermediate Q of methane monooxygenase. *Science* **275**, 515–518 (1997).
- Tinberg, C. E. & Lippard, S. J. Revisiting the mechanism of dioxygen activation in soluble methane monooxygenase from *M. capsulatus* (Bath): Evidence for a multi-step, proton-dependent reaction pathway. *Biochemistry* **48**, 12145–12158 (2009).
- Whittington, D. A. & Lippard, S. J. Crystal structures of the soluble methane monooxygenase hydroxylase from *Methylococcus capsulatus* (Bath) demonstrating geometrical variability at the dinuclear iron active site. *J. Am. Chem. Soc.* **123**, 827–838 (2001).
- Paulsen, K. E. *et al.* Oxidation-reduction potentials of the methane monooxygenase hydroxylase component from *Methylosinus trichosporium* OB3b. *Biochemistry* **33**, 713–722 (1994).
- Froland, W. A., Andersson, K. K., Lee, S. K., Liu, Y. & Lipscomb, J. D. Methane monooxygenase component B and reductase alter the regioselectivity of the hydroxylase component-catalyzed reactions. *J. Biol. Chem.* **267**, 17588–17597 (1992).
- Liu, Y., Nesheim, J. C., Lee, S. K. & Lipscomb, J. D. Kinetic and spectroscopic characterization of intermediates and component interactions in reactions of methane monooxygenase from *Methylococcus capsulatus* (Bath). *J. Am. Chem. Soc.* **117**, 10174–10185 (1995).
- Liu, Y., Nesheim, J. C., Lee, S. K. & Lipscomb, J. D. Gating effects of component B on oxygen activation by the methane monooxygenase hydroxylase component. *J. Biol. Chem.* **270**, 24662–24665 (1995).

15. Bailey, L. J., McCoy, J. G., Phillips, G. N. & Fox, B. G. Structural consequences of effector protein complex formation in a diiron hydroxylase. *Proc. Natl Acad. Sci. USA* **105**, 19194–19198 (2008).
16. Sazinsky, M. H. & Lippard, S. J. Correlating structure with function in bacterial multicomponent monooxygenases and related diiron proteins. *Acc. Chem. Res.* **39**, 558–566 (2006).
17. Sazinsky, M. H., Dunten, P. W., McCormick, M. S., DiDonato, A. & Lippard, S. J. X-ray structure of a hydroxylase-regulatory protein complex from a hydrocarbon-oxidizing multicomponent monooxygenase, *Pseudomonas* sp. OX1 phenol hydroxylase. *Biochemistry* **45**, 15392–15404 (2006).
18. Walters, K. J., Gassner, G. T., Lippard, S. J. & Wagner, G. Structure of the soluble methane monooxygenase regulatory protein B. *Proc. Natl Acad. Sci. USA* **96**, 7877–7882 (1999).
19. Brandstetter, H., Whittington, D. A., Lippard, S. J. & Frederick, C. A. Mutational and structural analyses of the regulatory protein B of soluble methane monooxygenase from *Methylococcus capsulatus* (Bath). *Chem. Biol.* **6**, 441–449 (1999).
20. Chang, S. L., Wallar, B. J., Lipscomb, J. D. & Mayo, K. H. Residues in *Methylosinus trichosporium* OB3b methane monooxygenase component B involved in molecular interactions with reduced- and oxidized-hydroxylase component: a role for the N-terminus. *Biochemistry* **40**, 9539–9551 (2001).
21. Blazyk, J. L., Gassner, G. T. & Lippard, S. J. Intermolecular electron-transfer reactions in soluble methane monooxygenase: A role for hysteresis in protein function. *J. Am. Chem. Soc.* **127**, 17364–17376 (2005).
22. McCormick, M. S. & Lippard, S. J. Analysis of substrate access to active sites in bacterial multicomponent monooxygenase hydroxylases: X-ray crystal structure of xenon-pressurized phenol hydroxylase from *Pseudomonas* sp OX1. *Biochemistry* **50**, 11058–11069 (2011).
23. Lee, S. K. & Lipscomb, J. D. Oxygen activation catalyzed by methane monooxygenase hydroxylase component: proton delivery during the O–O bond cleavage steps. *Biochemistry* **38**, 4423–4432 (1999).
24. Song, W. J. *et al.* Active site threonine facilitates proton transfer during dioxygen activation at the diiron center of toluene/o-xylene monooxygenase hydroxylase. *J. Am. Chem. Soc.* **132**, 13582–13585 (2010).
25. Rardin, R. L., Tolman, W. B. & Lippard, S. J. Monodentate carboxylate complexes and the carboxylate shift: Implications for polymetalloprotein structure and function. *New J. Chem.* **15**, 417–430 (1991).
26. Whittington, D. A., Rosenzweig, A. C., Frederick, C. A. & Lippard, S. J. Xenon and halogenated alkanes track putative substrate binding cavities in the soluble methane monooxygenase hydroxylase. *Biochemistry* **40**, 3476–3482 (2001).
27. Whittington, D. A., Sazinsky, M. H. & Lippard, S. J. X-ray crystal structure of alcohol products bound at the active site of soluble methane monooxygenase hydroxylase. *J. Am. Chem. Soc.* **123**, 1794–1795 (2001).
28. Rosenzweig, A. C. *et al.* Crystal structures of the methane monooxygenase hydroxylase from *Methylococcus capsulatus* (Bath): Implications for substrate gating and component interactions. *Proteins* **29**, 141–152 (1997).
29. Coufal, D. E. *et al.* Sequencing and analysis of the *Methylococcus capsulatus* (Bath) soluble methane monooxygenase genes. *Eur. J. Biochem.* **267**, 2174–2185 (2000).
30. Valentine, A. M., Stahl, S. S. & Lippard, S. J. Mechanistic studies of the reaction of reduced methane monooxygenase hydroxylase with dioxygen and substrates. *J. Am. Chem. Soc.* **121**, 3876–3887 (1999).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grant GM 32114 from the National Institute of General Medical Sciences to S. J. Lippard. We thank the staff at the Advanced Light Source beamline 8.2.2. in Lawrence Berkeley National Laboratory for the data collection, S. C. Harrison for resources and comments on the manuscript, and T. C. Johnstone, A. D. Liang and T.-T. Lu for discussions.

Author Contributions S. J. Lee designed experiments, purified proteins and measured the enzyme activity, analysed data, and wrote the manuscript; M.S.M. analysed the enzyme cavity, analysed data, prepared figures, and wrote the manuscript; S. J. Lippard directed the project, designed experiments, analysed data, and wrote the manuscript; and U.-S.C. obtained crystals, solved and refined the structures, analysed data, and wrote the manuscript. All authors discussed the results and commented on the manuscripts.

Author Information Atomic coordinates and structure factors for the crystal structure of the H–B complex have been deposited with the Protein Data Bank under the accession code 4GAM. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. J. Lippard (lippard@mit.edu) or U.-S.C. (uhnsoo@med.umich.edu).

METHODS

Soluble MMO fermentation and purification of MMOH. *M. capsulatus* (Bath) cultures were fermented and MMOH was purified by DEAE-sepharose fast-flow, S-300 size-exclusion, Q sepharose, and S-200 size-exclusion chromatography^{9,29–32}. The final eluent was concentrated to form a pale yellow solution.

MMOB and truncated version of MMOB expression and purification. The wild-type and truncated versions of MMOB were prepared recombinantly in *Escherichia coli*. From a recombinant glycerol stock of native MMOB (pkk223-3-mmoB, JM105) or truncated versions of MMOB (pET22b(+)-mmoB, BL21(DE3)), cells were grown and expressed for 3 h at 37 °C. The native and truncated regulatory proteins were purified using Q sepharose fast-flow and S-75 size-exclusion chromatography to obtain a colourless solution^{18,19,29–31}.

Enzyme activity measurement of MMOH in the presence of the full-length or N-terminal truncated regulatory subunit (MMOB). MMOH (1.0 µM), MMOB (2.0 µM) and MMOR (0.5 µM) were incubated with propylene in 25.0 mM phosphate buffer at pH 7.0. Steady-state kinetic data were recorded by using an HP8452 diode array spectrophotometer^{19,31}. The temperature was controlled at 45 °C with a circulating water bath. The reaction was initiated by the addition of NADH (167.0 µM) in the presence of propylene (approximately 1.0 mM). The consumption of NADH was monitored spectrophotometrically at 340 nm and quantified by using an extinction coefficient of 6,220 M^{−1} cm^{−1}.

Crystallization, data collection and structure determination. Purified MMOH (α₂β₂γ₂) and MMOB, which was stored in 30 mM HEPES, pH 7.5, 100 mM NaCl and 1 mM TCEP, were mixed at a 1:2.2 molar ratio and the final concentration was adjusted to ~10 mg ml^{−1}. Crystals were grown for 1 month at 18 °C by the sitting drop vapour diffusion method in 0.1 M MES, pH 6.5, and 15% PEG 20,000 (w/v). Crystals were flash-frozen in liquid nitrogen after transferring to a cryo-protectant

solution containing the precipitant and 20% glycerol. The crystal in space group *P*2₁2₁2₁, *a* = 183.6, *b* = 249.0, *c* = 122.3, gave recordable diffraction to a minimum Bragg spacing of 2.9 Å at the Advanced Light Source beamline 8.2.2. at the Lawrence Berkeley National Laboratory. Data were processed using XDS³³ and scaled with SCALA³⁴. Molecular replacement computations with MMOH_{ox} (PDB accession 1MTY) and MMOB (PDB accession 1CKV) were performed using the program Phaser³⁵. Model building and refinement were accomplished using Coot³⁶ and PHENIX.refine³⁷. We generated restraints for iron atoms and ligands using the program PHENIX (PHENIX.metal_coordination)³⁷ and applied them during the refinement of the diiron centre; no NCS restraints were applied in any stage of the refinement. The final refined model contains MMOH α-subunit (residues 15–526), β-subunit (residues 2–389) and γ-subunit (residues 3–168), and MMOB (residues 2–133) with *R*_{factor} and *R*_{free} values of 20.6 and 25.8, respectively.

31. Beauvais, L. G. & Lippard, S. J. Reactions of the peroxo intermediate of soluble methane monooxygenase hydroxylase with ethers. *J. Am. Chem. Soc.* **127**, 7370–7378 (2005).
32. Liu, K. E., Johnson, C. C., Newcomb, M. & Lippard, S. J. Radical clock substrate probes and kinetic isotope effect studies of the hydroxylation of hydrocarbons by methane monooxygenase. *J. Am. Chem. Soc.* **115**, 939–947 (1993).
33. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
34. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
35. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
37. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome

Yaser Hashem^{1*}, Amedee des Georges^{1*}, Jie Fu², Sarah N. Buss³, Fabrice Jossinet⁴, Amy Jobe⁵, Qin Zhang⁶, Hstau Y. Liao², Robert A. Grassucci¹, Chandrajit Bajaj⁶, Eric Westhof⁴, Susan Madison-Antenucci³ & Joachim Frank^{1,2,5}

Ribosomes, the protein factories of living cells, translate genetic information carried by messenger RNAs into proteins, and are thus involved in virtually all aspects of cellular development and maintenance. The few available structures of the eukaryotic ribosome^{1–6} reveal that it is more complex than its prokaryotic counterpart^{7,8}, owing mainly to the presence of eukaryote-specific ribosomal proteins and additional ribosomal RNA insertions, called expansion segments⁹. The structures also differ among species, partly in the size and arrangement of these expansion segments. Such differences are extreme in kinetoplastids, unicellular eukaryotic parasites often infectious to humans. Here we present a high-resolution cryo-electron microscopy structure of the ribosome of *Trypanosoma brucei*, the parasite that is transmitted by the tsetse fly and that causes African sleeping sickness. The atomic model reveals the unique features of this ribosome, characterized mainly by the presence of unusually large expansion segments and ribosomal-protein extensions leading to the formation of four additional inter-subunit bridges. We also find additional rRNA insertions, including one large rRNA domain that is not found in other eukaryotes. Furthermore, the structure reveals the five cleavage sites of the kinetoplastid large ribosomal subunit (LSU) rRNA chain, which is known to be cleaved uniquely into six pieces^{10–12}, and suggests that the cleavage is important for the maintenance of the *T. brucei* ribosome in the observed structure. We discuss several possible implications of the large rRNA expansion segments for the translation-regulation process. The structure could serve as a basis for future experiments aimed at understanding the functional importance of these kinetoplastid-specific ribosomal features in protein-translation regulation, an essential step towards finding effective and safe kinetoplastid-specific drugs.

The uniqueness of the *T. brucei* ribosomes, and those of other kinetoplastids, was highlighted by the discovery that the LSU rRNA chain is processed into six pieces^{10–12}. The rRNA secondary structure of *T. brucei* (Supplementary Fig. 1a, b) pointed to the existence of several interesting features, such as the size of several expansion segments. A comparison of *T. brucei* ribosomal protein sequences with those of eukaryotic homologues¹³ revealed the presence of *T. brucei*-specific large protein extensions. However, the arrangement and the structure of these features have remained unknown, although a low-resolution cryo-electron microscopy (cryo-EM) study of the *Trypanosoma cruzi* ribosome has been reported previously¹⁴. It is worth noting that gene expression in trypanosomes is primarily regulated post-transcriptionally (see ref. 15), at the protein-translation level. Several observations suggest that there are differences in both translation and translational regulation between *T. brucei* (and kinetoplastids in general) and other eukaryotes. For example, kinetoplastid messenger RNAs have a 35-nucleotide conserved spliced leader that is trans-spliced onto their 5' ends (see ref. 16) and that has a markedly modified cap. In addition, unique protein factors are involved in the maturation of the *T. brucei*

ribosomes¹⁷, whereas several eukaryotic initiation factors (eIF-4B, eIF-3 subunit P135 and eIF-3 subunit 3) and the eukaryotic ribosomal recycling factor (eRRF) are absent¹⁸.

Here we present a high-resolution cryo-EM structure of the *T. brucei* ribosome, and with this information we model the atomic structure of the ribosome using a combination of map segmentation^{19,20}, homology modelling, *ab initio* rRNA modelling based on covariation analysis with S2S and Assemble^{21,22}, and the molecular dynamics flexible fitting (MDFF) method²³ (see Methods).

The cryo-EM map was reconstructed from approximately 164,000 particles (see Methods). The *T. brucei* ribosome is in the inter-subunit-rotated state, with a transfer RNA at the exit (E) site (Fig. 1a, b). A measurement of local resolution (see Methods) shows that this is approximately 5 Å in most parts of the map (Supplementary Fig. 2a) but that there are small regions of high variance where it reaches 9 Å. Consistent with this, the three-dimensional variance²⁴ of the *T. brucei* ribosome map (Supplementary Fig. 2b) shows little structural variability, with the exception of the inter-subunit space where tRNA and translation-factor binding takes place, and areas around the L1 and P stalks. At the tRNA binding sites, the three-dimensional variance points mostly to the intermittent presence of tRNAs at the A and P sites. As our ribosomes are purified directly from *T. brucei* cell extracts, and therefore include a large proportion of translating ribosomes, such heterogeneity is expected. Data classification (see Methods) was attempted on a subset of the particles, but presumably because of the local nature of the variability it was not possible to identify distinct classes. Most of the *T. brucei* ribosome particles seem to exist in the same state. Periodically arranged densities for the phosphate groups along the rRNA strands and indications of protein side chains are visible (Supplementary Fig. 2c) in most regions of the map, and a comparison of the experimental cryo-EM map with simulated density maps generated from the atomic model, at a resolution of 5.0 Å (Supplementary Fig. 2d), validate our resolution estimation.

The first notable observation is the unusual size of several expansion segments (Fig. 1). In the small ribosomal subunit (SSU), expansion segments ES3S, ES6S, ES7S and ES9S seem to be several times larger than in other 80S ribosomes of known structure (Fig. 1c). Expansion segments ES4L, ES7L, ES19L, ES27L and ES31L of the LSU (Fig. 1d) are also considerably larger than in other known eukaryotic ribosomes. An additional domain on the LSU, which we refer to as the kinetoplastid-specific domain (KSD), is situated in the vicinity of ES7L (Fig. 1, red). These observations are consistent with the rRNA sequence and secondary structure in both ribosomal subunits²⁵ (Supplementary Fig. 1). We were able to model all proteins of the *T. brucei* ribosome (Fig. 2 and Supplementary Fig. 2d), except for some unstructured tails, and we were able to model the rRNAs in their entirety, with the exception of parts of ES27L and ES9S, which have poorly resolved densities due to their dynamic behaviour.

¹Howard Hughes Medical Institute (HHMI), Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA. ²Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA. ³Division of Infectious Diseases, Wadsworth Center, New York State Department of Health, Albany, New York 12201, USA. ⁴Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire (CNRS), Strasbourg 67084, France. ⁵Department of Biological Sciences, Columbia University, New York, New York 10027, USA. ⁶Department of Computer Science, Institute for Computational Engineering and Sciences, University of Texas, Austin, Texas 78712, USA.

*These authors contributed equally to this work.

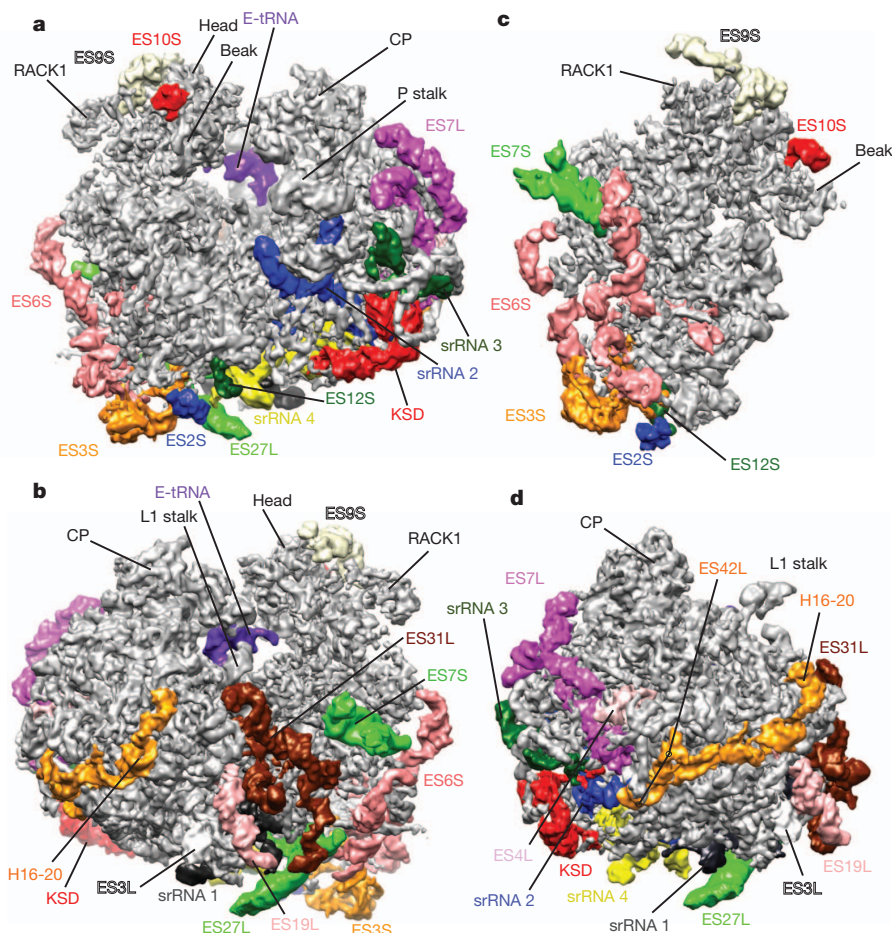


Figure 1 | High-resolution cryo-EM structure of the *T. brucei* ribosome. The density map was filtered with a locally varying bandpass, according to the local resolution measurements (see Methods and Supplementary Fig. 2A). In all panels, the largest rRNA expansion segments (ESs) are shown in different colours. **a**, The *T. brucei* ribosome viewed from the front (P stalk and beak side). **b**, The *T. brucei* ribosome viewed from the back (L1 stalk and platform side). **c**, The solvent side of the SSU. **d**, The solvent side of the LSU. CP, central protuberance; E-tRNA, exit-site RNA.

Our atomic model of the *T. brucei* ribosome reveals the structures of its expansion segments and its additional rRNA insertions (in both subunits) (Fig. 2a, b). These expansion segments and insertions represent a substantial additional mass compared to other known eukaryotic ribosomes (Fig. 2c, d). In spite of this addition and the change in the rRNA arrangement at several sites (Supplementary Fig. 3a, c), the structures and binding sites of ribosomal proteins are conserved (Supplementary Fig. 3b, d), as exemplified by proteins S7e on the SSU and L19e on the LSU (Fig. 3a). The high homology of these proteins with their counterparts in yeast contrasts with the differences in the arrangement of the surrounding rRNA context, formed by expansion segments ES6S, ES27L and ES31L (Fig. 3a).

The largest expansion segments are concentrated mostly in the regions of the LSU L1 stalk and the platform of the SSU. These large expansion segments form four additional inter-subunit bridges that are not seen in either of the two available 80S structures^{1–4} (Fig. 3a). We refer to these bridges as B_{Tb-1} B_{Tb-2} B_{Tb-3} and B_{Tb-4} (Tb, *T. brucei*). B_{Tb-1} is formed by three contacts (Fig. 3a and Supplementary Fig. 4a); B_{Tb-1a}, SSU ES7S helix A with LSU ES31L helix C. B_{Tb-1b}, SSU ES7S helix A with L34e carboxy-terminal extension accompanying ES31 helix C. B_{Tb-1c}, SSU ES6S helix F with LSU ES31L helix C. This first bridge seems to be stronger in the *T. cruzi* density map published previously¹⁴ (Supplementary Fig. 4a). B_{Tb-2} is formed by SSU ES6S helix G and LSU ES27L helix A (Fig. 3a and Supplementary Fig. 4b). This bridge is not established in *T. cruzi* despite the presence of its components. B_{Tb-3} is formed by SSU ES12S and a trypanosome-specific expansion segment grafted onto one of four short ribosomal RNAs in *T. brucei* (we refer to this here as srRNA-4-ES; expansion segment on short rRNA 4) (Fig. 3a and Supplementary Fig. 4c). This bridge is the strongest among the four, as seen from the *T. brucei* and

T. cruzi density maps. B_{Tb-4} is formed by SSU ES3S helix C and LSU ES27L helix B (Fig. 3a and Supplementary Fig. 4d), but the cryo-EM density of this bridge is poorly resolved and can be observed only at a low-density threshold in *T. brucei*, probably owing to the inherent flexibility of ES3S and ES27L. In contrast, in *T. cruzi* this bridge is visible at a relatively high-density threshold.

In terms of the enigmatic processing of LSU rRNA into six pieces, two large ribosomal RNAs, LSU- α and LSU- β , and four short ribosomal RNAs (which we refer to here as srRNA 1, srRNA 2, srRNA 3 and srRNA 4), our structure shows the arrangement of these pieces. Although the processing of the LSU rRNA into two large pieces has been observed in other organisms (see ref. 10), only trypanosomes and other kinetoplastids¹⁰ show such extensive processing. srRNA 1, which is composed of approximately 220 nucleotides^{11,12}, is generated from the main immature rRNA of LSU by the removal of two flanking sequences (Supplementary Fig. 5a), by splitting the main immature rRNA chain of the LSU into two components, LSU- α and LSU- β (Fig. 3a). These flanking sequences are localized at the positions of LSU ES20L and ES26L, respectively, which are also found in other eukaryotic ribosomes⁹ (next to H54 on LSU domain III)¹². It is possible that LSU expansion segments ES20L and ES26L in kinetoplastids have evolved to become cleavage sites.

srRNA 1 is structurally very similar to its homologous region on domain III in yeast (Supplementary Fig. 5b), except that it bears an expansion segment, which we name srRNA-1-ES. It is interesting to note that LSU ES19L is two to three times larger in *T. brucei* than it is in other 80S ribosomes (Fig. 3). srRNA 1 cleavage in *T. brucei* seems to remove a piece of rRNA that would otherwise collide with ES19L. It is possible that this cleavage was an expedient evolutionary solution to allow room for the unusually large ES19L. However, the functional importance of ES19L is unknown.

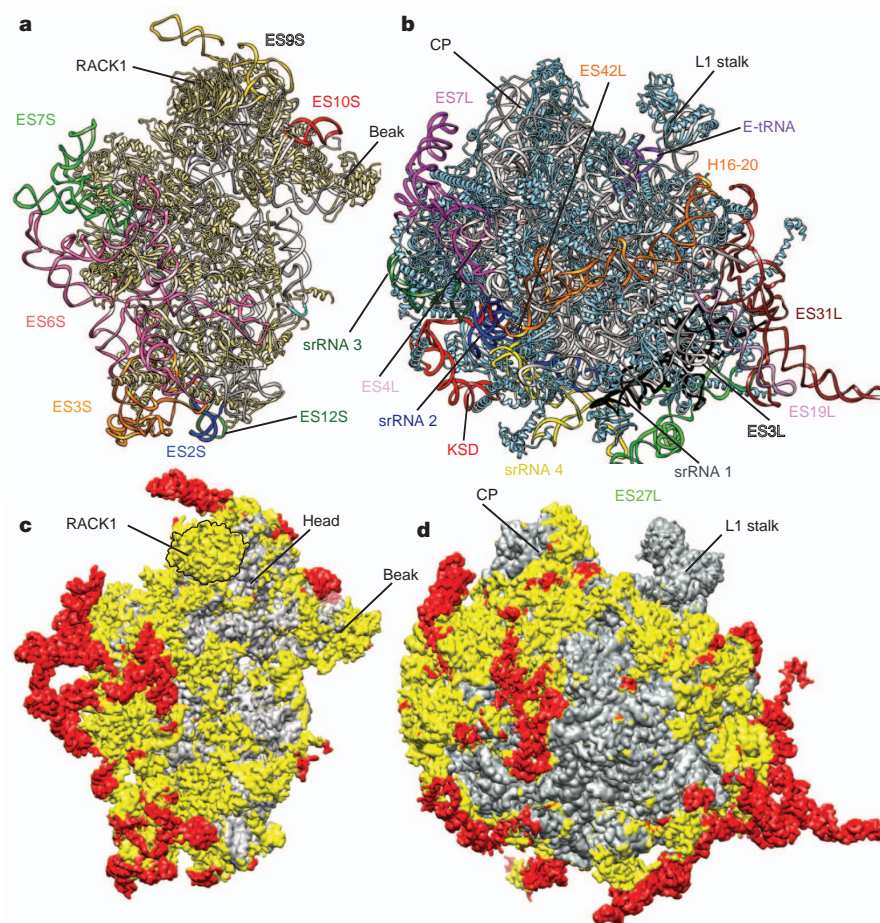


Figure 2 | Atomic model of the *T. brucei* ribosome. **a, b,** Atomic models of SSU (**a**) and LSU (**b**) with the expansion segments coloured and annotated as in Fig. 1. **c, d,** Atomic models of SSU (**c**) and LSU (**d**) in surface presentations. Grey regions indicate the location of conserved common core elements of all (prokaryotic and eukaryotic) ribosomes. Yellow regions highlight eukaryote-specific conserved elements, including those for trypanosomes. Red regions indicate trypanosome-specific elements that do not exist in other 80S ribosomes.

The remaining srRNAs (srRNA 2, srRNA 3 and srRNA 4 (180, 70 and 140 nucleotides, respectively)) were suggested previously, based on sequence conservation and two-dimensional modelling, to interact with one another and mimic the missing last domain of the LSU- β rRNA in the *T. brucei* ribosome, corresponding to domain VI in yeast and other 80S ribosomes¹². Our structure confirms this hypothesis and shows, in addition, that srRNA 2, srRNA 3 and srRNA 4 interact directly with the KSD (Fig. 4a, left and middle) to form one single stable domain similar to domain VI in other 80S ribosomes (Fig. 4a, right). Similar to the processing of srRNA 1, the processing of srRNA 2, srRNA 3 and srRNA 4 may be, at least in part, a structural necessity. Indeed, the existence of KSD in the structure and conformation observed is possible only if several pieces of rRNA are cleaved, as they would otherwise collide with KSD. The cleavage sites correspond to LSU expansion segments ES36L and part of ES39L in other eukaryotic ribosomes, and the cleavage yields the three srRNAs. The bundle formed by srRNA 2, srRNA 3, srRNA 4 and KSD is anchored in place through two specific protein extensions on L14e and L31e that hug the rRNA bundle (Fig. 4a, middle). There is also a unique interaction between srRNA 2 and domain I of LSU- α rRNA (Fig. 4a, arrow). This interaction occurs between a proximal stem-loop on srRNA 2 and another stem-loop, part of a *T. brucei* expansion segment between helices 19 and 20 on the LSU- α rRNA domain I, which we call ES42L. Although the effect of this connection is unknown, we speculate that it may contribute to maintaining the structure of the srRNAs bundle. *T. brucei* ribosome 5S and 5.8S rRNAs are very similar to their counterparts in yeast, except that ES3L on the 5.8S rRNA is larger.

No trypanosome-specific proteins were found, based on the analysis of the density map. However, several *T. brucei* ribosomal proteins have evolved specific helical extensions that interact with the additional rRNA mass and contribute towards stabilizing the extended expansion segments in their observed conformations (Fig. 4b). In other cases,

ribosomal protein tails have adopted different structures. On the LSU, L4 presents a long amino-terminal helical tail, different in the other available structures, which is probably due to the presence of a large, trypanosome-specific rRNA helical insertion in LSU ES7L, helix A (Fig. 4b). On the SSU, the trypanosomal RACK1 protein (also known as guanine nucleotide-binding protein subunit β -like protein) is found in its conserved binding site (Fig. 2a), which contrasts with *T. cruzi* ribosome¹⁴, in which trypanosomal RACK1 was missing from the density map, for a reason that is unresolved²⁶.

An interesting protein-rRNA interaction is formed by ribosomal protein S6e with expansion segments ES3S and ES6S (Fig. 3a, black circle). Ribosomal protein S6e possess an unstructured C-terminal tail, unresolved in both yeast 80S^{1,2} and *Tetrahymena thermophila* 40S subunit structures⁵. This tail is involved in translation regulation by phosphorylation of several serine residues recognized by S6e kinases (RSK and S6K) in higher eukaryotes and many lower eukaryotes (see ref. 27). The *T. brucei* S6e C-terminal tail contains a conserved RXXSS motif, recognizable by S6 kinases, and seems to have a helical structure (Fig. 3a, black circle, and Supplementary Fig. 6). The S6eC-terminal tail is surrounded by additional parts of SSU expansion segments ES3S and ES6S shielding the phosphorylation site, Ser 233 and Ser 234 (Supplementary Fig. 6), and could render it inaccessible to kinases on the ribosome. This is suggestive of a difference in translation regulation between trypanosomes and higher eukaryotes.

The role of the kinetoplastid-specific expansion segments in the regulation of the translation process remains unknown. In fact, very little is known about the translation regulation in kinetoplastids. However, the localization and the arrangement of the specific expansion segments, revealed by our structure, suggest the importance of several elements. In *T. brucei*, SSU expansion segments ES6S and ES7S are localized near the mRNA channel exit and are poised for a potential role in the regulation of translation initiation. The localization of

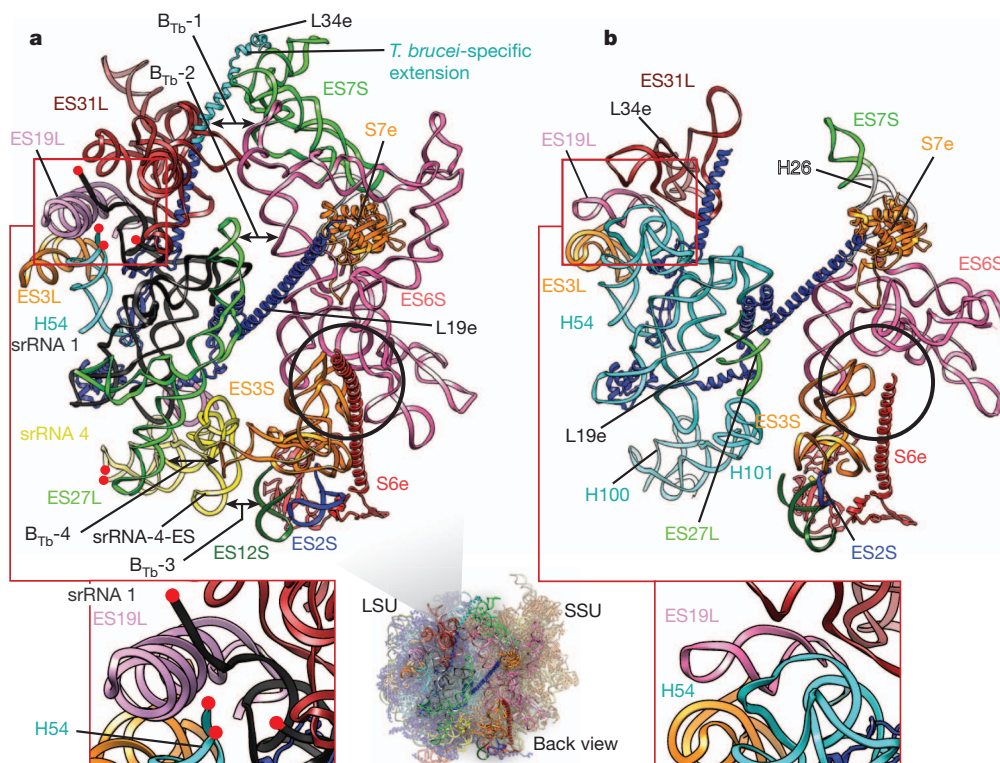


Figure 3 | Comparison between *T. brucei* and yeast ribosomes.
a, b, Back view of rRNA expansion segments at the back of the *T. brucei* (a) and yeast (b) ribosomes. *T. brucei*-specific inter-subunit bridges are indicated by double arrows in a. Insets show enlargements of the srRNA 1 cleavage site in *T. brucei* and its counterpart region in yeast. Black circles highlight the interaction of S6e with ES3S and ES6S in both ribosomes. Red spheres in a denote srRNA 1 and srRNA 4 cleaved ends.

expansion segments ES6S and ES7S on the SSU, near the mRNA channel exit, suggests their possible interaction with the mRNA spliced leader and its unusually modified cap, but also with eIF3 (refs 28, 29). It is interesting to note that SSU expansion segments ES6S and ES7S collide with the proposed binding site of eIF3 (ref. 29) (Fig. 4c). The initiation factor eIF3 does exist in trypanosomes¹⁸ (Supplementary Fig. 7); in *T. brucei* it contains at least seven or eight subunits. The structure here suggests that the proposed binding site of eIF3 (ref. 29) may require reconsideration. The *T. brucei* ribosome structure may

alternatively suggest that the SSU interaction with eIF3 is different from that in other eukaryotes, and that expansion segments ES6S and ES7S may have an important role in this interaction.

It was found that the *T. brucei* ribosome interacts with specific factors involved in translation regulation (such as L34 and L37 (not to be confused with the LSU proteins of the same name))¹⁷ that are required in the maturation and cytoplasmic exportation of the LSU. It is possible that kinetoplastid-specific expansion segments are involved in the binding of such specific factors that regulate the translation process. This hypothesis

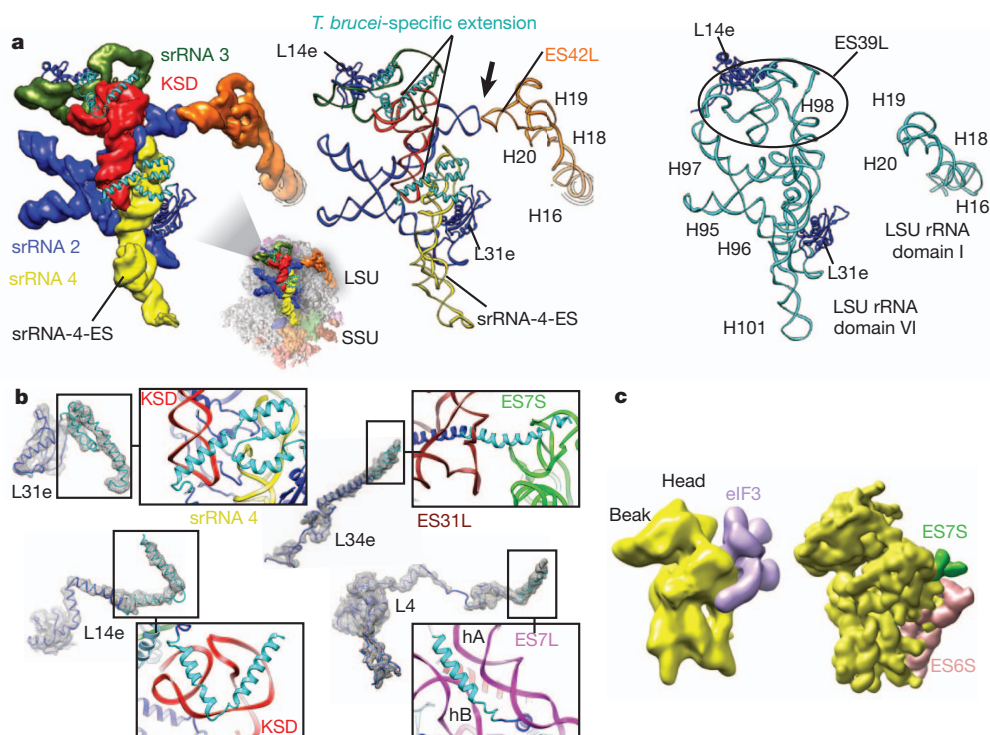


Figure 4 | *T. brucei* srRNAs and protein extensions. **a,** Left and middle, srRNA 2, srRNA 3, srRNA 4 and KSD cryo-EM density segmentations with their atomic models. Right, yeast LSU rRNA domain I and LSU rRNA domain VI. Arrow shows the kissing-loop interaction. Black circle surrounds ES39L in the yeast ribosome. **b,** Ribosomal proteins presenting specific extensions in *T. brucei* (cyan ribbon). Insets show enlargements of the interactions of their extensions with the surrounding rRNA. hA, helix A, hB, helix B. **c,** Left, cryo-EM-derived model of the eIF3-binding site (recreated according to ref. 21). Right, segmented map of the SSU from the *T. brucei* ribosome, filtered to 20 Å.

is supported by the observation that *T. brucei* expansion segments contain several pyrimidine-rich internal loops, that have been suggested to be important for protein recognition or binding to other RNA elements³⁰.

In conclusion, our cryo-EM study reveals that the *T. brucei* ribosome has an unusual structure and shows the unique arrangement of its expansion segments. These features may have consequences for the regulation of translation in trypanosomes. Our results may explain, in structural terms, the rationale for the processing of LSU rRNA into six pieces. More generally, our structure highlights the extraordinary diversity of ribosome structures among eukaryotes. We believe that these findings provide a basis for future investigations of translation regulation in kinetoplastids and of the function of the specific expansion segments, opening the door to the development of new drugs.

METHODS SUMMARY

Trypanosome culture and ribosome isolation. Pro-cyclic forms of *T. brucei* strain TREU 667 were grown at 27 °C. Cell pellets were washed and resuspended in SKS buffer containing 5 mM EGTA, 0.5% Triton X-100 and protease inhibitor cocktail (0.5 mM TLCK (N-tosyl-L-leucine chloromethyl ketone), 0.1 mM PMSF (phenylmethylsulphonyl fluoride), 25 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ leupeptin). Cells were freeze-thawed. Several centrifugation steps were used to obtain a ribosome-enriched pellet that was resuspended in a buffer containing 20 mM HEPES, pH 7.2, 10 mM MgCl₂, 500 mM KCl and 5 mM β-mercaptoethanol. The suspension was clarified and ribosomes were purified.

Electron microscopy and image processing. Three microlitres of each sample were applied to holey carbon grids (carbon-coated Quantifoil 2/4 grid, Quantifoil Micro Tools, GmbH) containing an additional continuous thin layer of carbon, and glow-discharged. Grids were blotted for 6 s at 4 °C in 100% humidity and vitrified by plunging rapidly into liquid ethane at -180 °C with a Vitrobot (FEI). Cryo-EM data were acquired under low-dose conditions (25 electrons (e⁻) per Å²) on a Tecnai F30 Polara (FEI) operating at 300 kV. The micrographs were taken at a nominal magnification of ×59,000 on Kodak SO-163 film. The data were processed using SPIDER. The data windows were inspected manually. The contrast transfer function (CTF) of the reconstructions was corrected by phase-flipping the particles using the defocus value estimated for each micrograph and a single reconstruction was obtained from the entire data set using conjugate gradients with regularization (BP CG command in SPIDER).

Ribosomal RNA and protein modelling and fitting. The *T. brucei* SSU and LSU rRNA were modelled by homology, and rRNA expansion segments were modelled mostly *ab initio* by performing a covariation analysis against a number of orthologous rRNA sequences found in other related kinetoplastids. The ribosomal proteins were modelled mostly by homology. The resulting models were fitted using the MDFF method²³.

Full Methods and any associated references are available in the online version of the paper.

Received 3 October; accepted 21 December 2012.

Published online 10 February 2013.

- Ben-Shem, A. *et al.* Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203–1209 (2010).
- Ben-Shem, A. *et al.* The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**, 1524–1529 (2011).
- Armache, J.P. *et al.* Localization of eukaryote-specific ribosomal proteins in a 5.5 Å cryo-EM map of the 80S eukaryotic ribosome. *Proc. Natl Acad. Sci. USA* **107**, 19754–19759 (2010).
- Armache, J.P. *et al.* Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5 Å resolution. *Proc. Natl Acad. Sci. USA* **107**, 19748–19753 (2010).
- Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730–736 (2011).
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S. & Ban, N. Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334**, 941–948 (2011).
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M. & Ban, N. Atomic structures of the eukaryotic ribosome. *Trends Biochem. Sci.* **37**, 189–198 (2012).
- Wilson, D. N. & Doudna, C. J. H. The structure and function of the eukaryotic ribosome. *Cold Spring Harb. Perspect. Biol.* **4**, a011536 (2012).
- Yokoyama, T. & Suzuki, T. Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments. *Nucleic Acids Res.* **36**, 3539–3551 (2008).

- White, T. C., Rudenko, G. & Borst, P. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNA. *Nucleic Acids Res.* **14**, 9471–9489 (1986).
- Cordingley, J. S. & Turner, M. J. 6.5 S RNA: preliminary characterization of unusual small RNAs in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **1**, 91–96 (1980).
- Campbell, D. A., Kubo, K., Graham Clark, C. & Boothroyd, J. C. Precise identification of cleavage sites involved in the unusual processing of trypanosome ribosomal RNA. *J. Mol. Biol.* **196**, 113–124 (1987).
- Berriman, M. *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422 (2005).
- Gao, H., Juri Ayub, M., Levin, M. L. & Frank, J. The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. *Proc. Natl Acad. Sci. USA* **102**, 10206–10211 (2005).
- Clayton, C. & Shapiro, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol. Biochem. Parasitol.* **156**, 93–101 (2007).
- Michaeli, S. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol.* **6**, 459–474 (2011).
- Ciganda, M. & Williams, N. Characterization of a novel association between two trypanosome-specific proteins and 5S rRNA. *PLoS ONE* **7**, e30029 (2012).
- Ivens, A. C. *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436–442 (2005).
- Zhang, Q., Bettadapura, R. & Bajaj, C. Macromolecular structure modeling from 3D EM using VolRover 2.0. *Biopolymers* **97**, 709–731 (2012).
- Pintilie, G., Zhang, J., Goddard, T., Chiu, W. & Gossard, D. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 429–438 (2010).
- Jossinet, F. & Westhof, E. Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* **21**, 3320–3321 (2005).
- Jossinet, F., Ludwig, T. E. & Westhof, E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057–2059 (2010).
- Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683 (2008).
- Liao, H. Y. & Frank, J. Classification by bootstrapping in single particle methods. *IEEE Int. Symp. Biom. Imaging* 169–172 (2010).
- Cannone, J. J. *et al.* The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *Bioinformatics* **3**, 2 (2002).
- Ayub, M. J., Atwood, J., Nuccio, A., Tarleton, R. & Levin, M. J. Proteomic analysis of the *Trypanosoma cruzi* ribosomal proteins. *Biochem. Biophys. Res. Commun.* **382**, 30–34 (2009).
- Meyhuas, O. Physiological roles of ribosomal protein S6: One of its kind. *Int. Rev. Cell Mol. Biol.* **268**, 1–37 (2008).
- Srivastava, S., Verschoor, A. & Frank, J. Eukaryotic initiation factor 3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface. *J. Mol. Biol.* **226**, 301–304 (1992).
- Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A. & Nogales, E. structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310**, 1513–1515 (2005).
- Lescrier, E. M. H. P. *et al.* Structure of the pyrimidine-rich internal loop in the poliovirus 30'-UTR: The importance of maintaining pseudo-2-fold symmetry in RNA helices containing two adjacent non-canonical base-pairs. *J. Mol. Biol.* **331**, 759–769 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is dedicated to the memory of Mariano Levin, who collaborated with J.F. and S.M.A. on the ribosomes from *T. cruzi* and *T. brucei*. We thank G. Cardone for assistance in the local resolution computation, and M. Thomas for her assistance with the preparation of figures. We wish to thank N. Williams for an useful discussion about the *T. brucei* LSU rRNA processing. This work was supported by the Howard Hughes Medical Institute (HHMI) and the National Institutes of Health (NIH) R01 GM29169 (to J.F.), L'Agence Nationale de la recherche (ANR) project AMIS ARN ANR-09-BLAN-0160 (E.W. and F.J.), as well as NIH R01-EB004873 and R01-GM074258 (to Q.Z. and C.B.). S.N.B. was supported by a Centers for Disease Control (CDC) Emerging Infectious Diseases (EID) fellowship program.

Author Contributions Y.H., A.d.G., S.N.B., F.J., Q.Z., C.B., S.M.-A., E.W. and J.F. interpreted the data and wrote the manuscript. S.N.B. purified the *T. brucei* ribosomes. Y.H., J. Fu and R.A.G. carried out the cryo-EM experiments. H.Y.L. performed the three-dimensional variance estimation. Y.H., A.J. and Q.Z. performed the density-map segmentations. Y.H., A.d.G., J. Fu, A.J. and H.Y.L. carried out the cryo-EM data processing. Y.H. and F.J. modelled the rRNA. Y.H. and Q.Z. modelled the ribosomal proteins. J.F. directed research.

Author Information The electron microscopy map has been deposited in the European Molecular Biology Laboratory (EMBL) European Bioinformatics Institute Electron Microscopy Data Bank (EMDB) under accession code EMD-2239. Coordinates of electron-microscopy-based model have been deposited in the RCSB Protein Data Bank under accession numbers 3ZEQ, 3ZEX, 3ZEY and 3ZF7. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F. (jf2192@columbia.edu).

METHODS

Trypanosome culture and ribosome isolation. Pro-cyclic forms of *T. brucei* strain TREU 667 were grown at 27 °C in a semi-defined medium supplemented with 10% heat-inactivated fetal bovine serum and 25 mM HEPES, as described previously³¹. For ribosomal isolation, cells were grown to a density of $0.5\text{--}1.2 \times 10^7$ cells per ml in a 41 volume, and collected by 10 min centrifugation at 6,000g.

Ribosomes were prepared using a method described previously³², but with slight modifications. In brief, cell pellets were washed three times with SKS buffer (0.25 M sucrose, 5 mM KCl) and resuspended in SKS buffer containing 5 mM EGTA, 0.5% Triton X-100 and protease inhibitor cocktail (0.5 mM TLCK, 0.1 mM PMSF, $25 \mu\text{g ml}^{-1}$ aprotinin, $10 \mu\text{g ml}^{-1}$ leupeptin). The cells were freeze-thawed five times and ruptured further using 10 strokes of a Dounce homogenizer. Centrifugation steps described previously¹ were used to obtain a ribosome-enriched pellet, and the pellet was resuspended in a buffer containing 20 mM HEPES, pH 7.2, 10 mM MgCl_2 , 500 mM KCl and 5 mM β -mercaptoethanol. The suspension was clarified and ribosomes were purified using methods described previously¹, however, 20 mM HEPES, pH 7.2, was substituted for Tris in all buffers to reduce ribosomal aggregation.

Electron microscopy. Three microlitres of each sample was applied to holey carbon grids (carbon-coated Quantifoil 2/4 grid, Quantifoil Micro Tools GmbH) containing an additional continuous thin layer of carbon³³, and glow-discharged. Grids were blotted for 6 s at 4 °C in 100% humidity and vitrified by plunging rapidly into liquid ethane at -180°C with a Vitrobot (FEI)^{34,35}. Data were acquired under low-dose conditions (25 e^- per \AA^2) on a Tecnai F30 Polara (FEI) operating at 300 kV. The micrographs were taken at a nominal magnification of $\times 59,000$ on Kodak SO-163 film using the automated data collection system AutoEMation³⁶, and scanned on a Nikon Super Coolsan 9000 ED with a step size of 6.35 μm , resulting in a pixel size of 1.09 \AA on the object scale.

Image processing. The data were processed using SPIDER³⁷. Approximately 700,000 particle windows were automatically extracted from 1,100 film-recorded micrographs using the lfpick.spi spider batch file⁸ (a batch file of SPIDER commands that extracts potential particles by template matching) with a window size of 500 pixels and a pixel size of 1.09 \AA . The windows were inspected manually; only particles with high contrast were selected, and those in contact with other particles were eliminated. This conservative manual verification step yielded approximately 164,000 particles in total. Reconstruction was carried out following standard SPIDER protocols for reference-based reconstruction³⁸, except that contrast transfer function of the reconstructions was corrected by phase-flipping the particles using the defocus value estimated for each micrograph and a single reconstruction was obtained from the entire data set using conjugate gradients with regularization (BP CG command in SPIDER). The resolution at FSC = 0.5 for the final reconstruction was estimated at 5.57 \AA (Supplementary Fig. 2C). However, this resolution is an average of the local resolutions over all of the ribosome and it underestimates most of the ribosome volume that is resolved at approximately 5.0 \AA (see below).

Maximum-likelihood classification³⁹, on a subset of 40,000 particles taken from the full data set, was attempted using the XMIPP package⁴⁰. The full *T. brucei* reconstruction was used as a seed for four classes. The resulting classes were all very similar, as expected based on the three-dimensional variance calculation (see below) and the resolution of the map, probably as a consequence of most of the ribosomes being in the same state. No further classification was attempted.

We attribute the high resolution obtained in our study to several factors; first, the CTF correction procedure by phase-flipping, which has the advantage of correcting the CTF more precisely at the micrograph level; second, carrying out the three-dimensional reconstructions (including all the phase-flipped particles) at once, instead of summing different reconstructions issued from different defocus groups; third, the global homogeneity of the sample in our study; and finally, the very careful particle selection.

Local resolution estimation and filtering. The local resolution was estimated on the final map using the 'blores' module, part of Bernard's Software Package (Bsoft; <http://lsbr.niams.nih.gov/bsoft/>)^{41,42}. The three-dimensional map depicting local resolutions showed that the ribosome volume has a resolution of approximately 5.0 \AA (at FSC = 0.5) that concerns mostly the core (Supplementary Fig. 2A), and a resolution between 5 and 6 \AA at the outer shell, whereas regions of higher variability have resolutions ranging between 7 and 9 \AA . The *T. brucei* cryo-EM map was filtered according to the estimated local resolution using the following procedure: the map was segmented in accordance with the local resolution using the Segger module⁴³ implemented in the Chimera UCSF program⁴⁴. The segments obtained were filtered according to their local resolution and then merged together into one map.

Map segmentation. A detailed description of the segmentation protocol can be found in the Supplementary Information section. A preliminary segmentation of the *T. brucei* map was carried out using Segger⁴³, now implemented in Chimera⁴⁴, and VolRover⁴⁵ (see also protocols described in refs 46–48). These preliminary

segments were used during the atomic modelling process. After the atomic model was obtained, we used it to re-segment the *T. brucei* map more accurately (see Supplementary Information).

Three-dimensional variance estimation. The particle windows were binned six times for the three-dimensional variance estimation step, to reduce the large amount of time and memory required for this operation, resulting in a pixel size of 6.54 \AA . The three-dimensional variance map was computed using the bootstrapping method, developed by several groups^{49–52}, in the following way: 40,000 bootstrap reconstructions were generated, each of which was reconstructed from n particle projections that were randomly sampled with replacement from the total set of approximately 164,000 particles. To enhance the signal-to-noise ratio of the three-dimensional variance, the bootstrap volumes were filtered to a resolution corresponding to approximately half the frequency of the first zero-crossing of the CTFs. The structural variance was estimated as the sample variance of the bootstrap volumes minus the variance of the noise, and the difference was then multiplied by n (ref. 50). In this estimation, the noise variance is assumed to be uniform across the map⁵⁰. Ideally, the three-dimensional variance map should reflect only regions of high conformational changes, including presence versus absence of a ligand such as a protein or a tRNA. Here, the low-pass filtration was necessary to reduce the noise variance. However, the filtering also introduces spatial correlation of the noise, and the resulting variance map is distorted to some extent by the noise variance (as verified using simulations).

Ribosomal RNA modelling. The *T. brucei* rRNAs were modelled using a workflow combining *ab initio* and homology approaches and can be found in the online Supplementary Information section, schematized in Supplementary Fig. 8. The 18S rRNA (SSU rRNA) was modelled based on the *T. thermophila* 40S crystal structures⁵. The LSU rRNA was modelled based on the yeast ribosome crystal structure². The LSU and SSU rRNAs were modelled by homology using the S2S tool²¹. *T. brucei* and other kinetoplastid rRNA sequences were fetched from the SILVA webserver (<http://www.arb-silva.de/>)⁵³ and the comparative RNA website and project (CRW site) (<http://www.rna.icmb.utexas.edu/>)⁵⁴. *T. brucei* expansion segments were modelled mostly *ab initio* by performing a covariation analysis on each one of these expansion segments sequences against a number of orthologous rRNA sequences found in other related kinetoplastids, using the mlocarna algorithm⁵⁵. This allowed deriving a consensus secondary structure that was used in the Assemble tool²² and three-dimensional models were generated for each expansion segment and fitted into the EM map.

Modelling of ribosomal proteins. The modelling workflow can be found in the Supplementary Information, and it is schematized in Supplementary Fig. 10. The ribosomal proteins were modelled by homology to mainly three crystal structures, the yeast ribosome², the *T. thermophila* 40S subunit in complex with eIF1 (ref. 5), and the 60S subunit with eIF6 (ref. 6). The *T. brucei* ribosomal protein sequences, used as inputs for the homology modelling, were taken from the National Center for Biotechnology Information (NCBI) protein databases (<http://www.ncbi.nlm.nih.gov/>). Swiss-Model web server was used for protein structure homology modelling (<http://swissmodel.expasy.org/>)^{56,57}. We used the Phyre and the Phyre2 web servers⁵⁸ for the *ab initio* modelling of the missing fragments and extensions into the density map. We assessed the structure of each model using the MolProbity web server (<http://molprobity.biochem.duke.edu/>)^{59,60}.

Molecular dynamics flexible fitting method. A detailed description of the MDFF method can be found in the Supplementary Information. The initial system was prepared for MDFF²³ using Visual Molecular Dynamics (VMD)⁶¹ and run in the Not Another Molecular Dynamics (NAMD) program⁶² for 1.5 ns of simulation time.

- Kaminsky, R., Beaudoin, E. & Cunningham, I. Cultivation of the life cycle stages of *Trypanosoma brucei* sspp. *Acta Trop.* **45**, 33–43 (1988).
- Gómez, E. B., Medina, G., Ballesta, J. P., Levin, M. J. & Téllez-Inón, M. T. Acidic ribosomal P proteins are phosphorylated in *Trypanosoma cruzi*. *Int. J. Parasitol.* **31**, 1032–1039 (2001).
- Grassucci, R. A., Taylor, D. J. & Frank, J. Preparation of macromolecular complexes for cryo-electron microscopy. *Nature Protocols* **2**, 3239–3246 (2007).
- Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).
- Wagenknecht, T., Frank, J., Boublik, M., Nurse, K. & Ofengand, J. Direct localization of the tRNA-anticodon interaction site on the *Escherichia coli* 30 S ribosomal subunit by electron microscopy and computerized image averaging. *J. Mol. Biol.* **203**, 753–760 (1988).
- Lei, J. & Frank, J. Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005).
- Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Shaikh, T. R. *et al.* SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nature Protocols* **3**, 1941–1974 (2008).

39. Scheres, S. H. W. *et al.* Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods* **4**, 27–29 (2007).
40. Scheres, S. H. W., Nuñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using Xmipp. *Nature Protocols* **3**, 977–990 (2008).
41. Heymann, J. B. Bsoft: image and molecular processing in electron microscopy. *J. Struct. Biol.* **133**, 156–169 (2001).
42. Heymann, J. B., Cardone, G., Winkler, D. C. & Steven, A. C. Computational resources for cryo-electron tomography in Bsoft. *J. Struct. Biol.* **161**, 232–242 (2008).
43. Pintilie, G., Zhang, J., Goddard, T., Chiu, W. & Gossard, D. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
44. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **13**, 1605–1612 (2004).
45. Baker, M. L., Yu, Z., Chiu, W. & Bajaj, C. Automated segmentation of molecular subunits in electron cryomicroscopy density maps. *J. Struct. Biol.* **156**, 432–441 (2006).
46. Yu, Z. & Bajaj, C. Automatic ultrastructure segmentation of reconstructed cryoEM maps of icosahedral viruses. *IEEE Trans. Image Process.* **14**, 1324–1337 (2005).
47. Zhang, Q., Bettadapura, R. & Bajaj, C. Macromolecular structure modeling from 3D EM using VolRover 2.0. *Biopolymers* **97**, 709–731 (2012).
48. Zeyen, Y. & Bajaj, C. Computational approaches for automatic structural analysis of large biomolecular complexes. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **5**, 568–582 (2008).
49. Penczek, P. A., Yang, C., Frank, J. & Spahn, C. M. Estimation of variance in single-particle reconstruction using the bootstrap technique. *J. Struct. Biol.* **154**, 168–183 (2006).
50. Zhang, W., Kimmel, M., Spahn, C. M. & Penczek, P. A. Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis. *Structure* **16**, 1770–1776 (2008).
51. Liao, H. Y. & Frank, J. Classification by bootstrapping in single particle methods. *IEEE Int. Symp. Biom. Imaging* 169–172 (2010).
52. Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416–420 (2008).
53. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
54. Cannone, J. J. *et al.* The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *Bioinformatics* **3**, 2 (2002).
55. Will, S., Joshi, T., Hofacker, I. L., Stadler, P. F. & Backofen, R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* **18**, 900–914 (2012).
56. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
57. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* **37**, D387–D392 (2009).
58. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
59. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
60. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
61. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph. Model.* **1**, 33–8–27–8 (1996).
62. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781 (2005).

CORRIGENDUM

doi:10.1038/nature11864

Corrigendum: The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma

Pedro A. Pérez-Mancera, Alistair G. Rust, Louise van der Weyden, Glen Kristiansen, Allen Li, Aaron L. Sarver, Kevin A. T. Silverstein, Robert Grützmann, Daniela Aust, Petra Rümmele, Thomas Knösel, Colin Herd, Derek L. Stemple, Ross Kettleborough, Jacqueline A. Brosnan, Ang Li, Richard Morgan, Spencer Knight, Jun Yu, Shane Stegeman, Lara S. Collier, Jelle J. ten Hoeve, Jeroen de Ridder, Alison P. Klein, Michael Goggins, Ralph H. Hruban, David K. Chang, Andrew V. Biankin, Sean M. Grimmond, Australian Pancreatic Cancer Genome Initiative, Lodewyk F. A. Wessels, Stephen A. Wood, Christine A. Iacobuzio-Donahue, Christian Pilarsky, David A. Largaespada, David J. Adams & David A. Tuveson

Nature **486**, 266–270 (2012); doi:10.1038/nature11114

In this Letter, several authors from the Australian Pancreatic Cancer Genome Initiative were inadvertently omitted: Christopher J. Scarlett and Warren Kaplan from the The Kinghorn Cancer Centre; and Aldo Scarpa from the ARC-NET Center for Applied Research on Cancer. Amitabha Das was incorrectly listed as Amithabad Das. Furthermore, the correct address for the ARC-NET Center for Applied Research on Cancer, University of Verona, is: Policlinico GB Rossi, Piazzale LA Scuro 10, 37134 Verona, Italy. In the acknowledgements section, S.A.W. was supported by the National Health and Medical Research Council of Australia (NHMRC); and the APGI investigators by the University of Verona and Italian Ministry of University and Research (FIRB RBAP10AHJB). These errors have been corrected in the HTML and PDF of the original paper.

CORRIGENDUM

doi:10.1038/nature11904

Corrigendum: Observations of increased tropical rainfall preceded by air passage over forests

D. V. Spracklen, S. R. Arnold & C. M. Taylor

Nature **489**, 282–285 (2012); doi:10.1038/nature11390

A. M. Makarieva, V. G. Gorshkov, D. Sheil, A. D. Nobre and P. Bunyard pointed out that the units in our Fig. 3b were confusing. To clarify, the legend should read: “**b**, Same as Fig. 2b, but for cumulative continental surface evaporation ($\sum ET$; here plotted per day of atmospheric transport) as a function of $\sum LAI$.” The HTML and PDF of the original paper have been corrected.

CORRIGENDUM

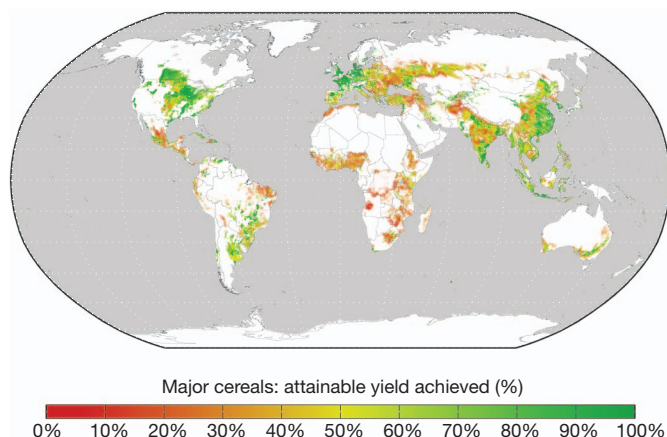
doi:10.1038/nature11907

Corrigendum: Closing yield gaps through nutrient and water management

Nathaniel D. Mueller, James S. Gerber, Matt Johnston, Deepak K. Ray, Navin Ramankutty & Jonathan A. Foley

Nature **490**, 254–257 (2012); doi:10.1038/nature11420

In this Letter, a minor error in the mapping program used to generate Fig. 1 incorrectly displayed reduced yield gap estimates across a few areas; this was most apparent in the Sahel region of Africa. The corrected Fig. 1 is shown below. This error does not affect the results or conclusion of the paper, and has been corrected in the HTML and PDF of the original paper.



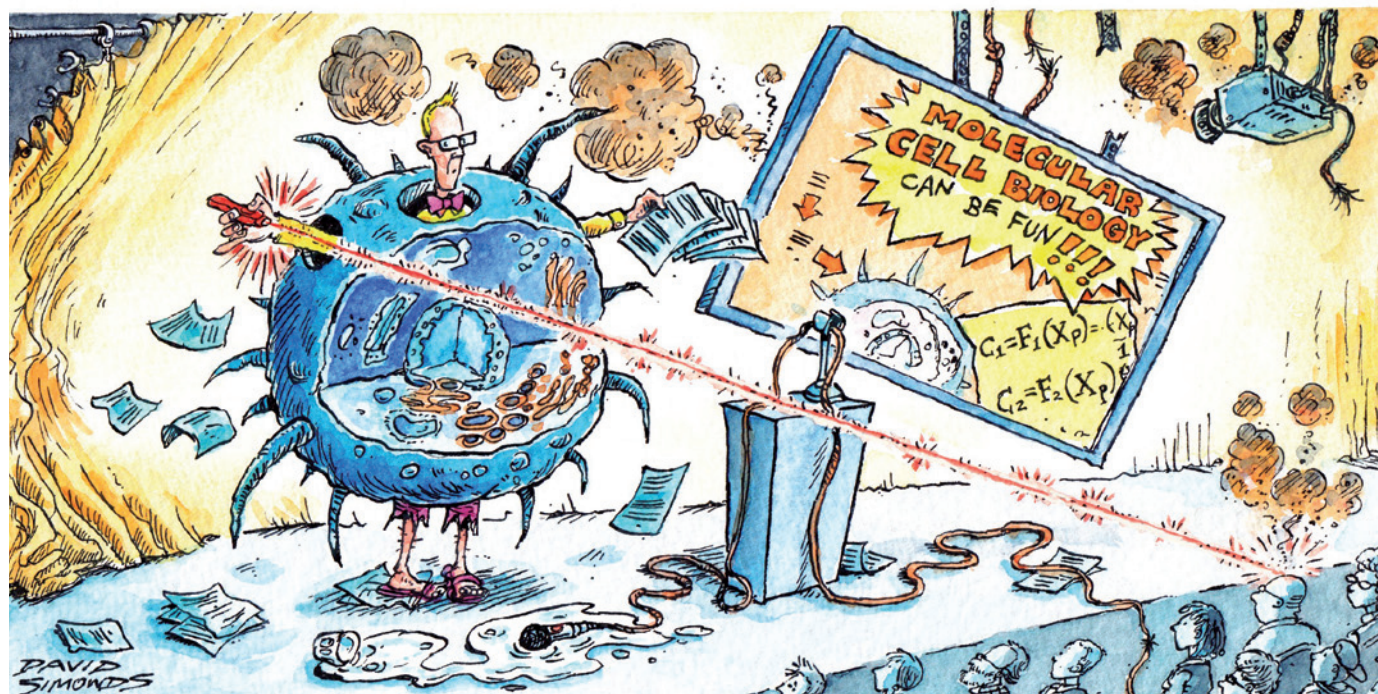
CAREERS

COLUMN The fossilization of the PhD harms students, employers and science **p.393**

CAREERS BLOG The latest discussions and science-careers issues go.nature.com/z8g4a7

NATUREJOBS For the latest career listings and advice www.naturejobs.com

DAVID SIMONDS



PRESENTATIONS

Pressure to perform

Talks offer scientists a chance to show off their work, but it is difficult to make an impact.

BY KAREN KAPLAN

It was 40 years ago, but thinking about one of his first talks to a large audience still makes Martin Raff shudder. Raff, a neurobiologist and now emeritus professor at the MRC Laboratory of Molecular Cell Biology at University College London, had been invited to present at a symposium on cell membranes. He had decided to rely on his slides for his 45-minute talk, forgoing notes or a crib sheet.

Raff walked onto the podium, faced the audience of several thousand people — and forgot everything he had planned to say. “As soon as I started, I hit a blank,” he recalls. “I couldn’t remember who I was talking to or why, or what my talk was about.” But Raff didn’t lose his cool: he asked the technician to put up the first slide, which helped him to regain his footing. “That gave me a clue, and I just ad-libbed to get myself into the introduction,” says Raff. “I don’t think the audience could tell what was going on, but

there were some embarrassing silences in the first minute or two.”

Many early-career scientists can relate to Raff’s panic. Plenty can — and does — go wrong with presentations. But a researcher can deliver a memorable talk by following a few basic principles, rehearsing often in front of different listeners and making a back-up plan. To excel at touting their work, scientists should become as familiar with the podium and microphone as they are with the lab bench.

SETTING THE STAGE

Before putting a talk together, presenters need to determine their audience’s level of expertise. Giving a talk is a balancing act: too technical, and some listeners get lost; too general, and they get bored. Talks of 30–45 minutes or longer tend to draw larger and more general audiences who need plenty of background and context. Shorter presentations are often for smaller groups that share the

speaker’s speciality, so these talks can be targeted, detailed and technical. One-size-fits-all presentations should be avoided, say veteran speakers. “If there’s any doubt, if you have any questions about the level you should be pitching to, you should consult the scientific organizing committee of the conference,” says Michael De Robertis, an astronomer at York University in Toronto, Canada. “That is critical. There are no excuses for getting it wrong.”

For larger audiences and longer talks, veteran presenters suggest that the speaker start by piquing the audience’s interest with a general but compelling question about the topic, followed by a discussion of why it is interesting, and a description of the presenter’s research question or hypothesis. “If you’re going to talk about cell-size control, you might say, ‘Why is it that we as humans grow to be so much larger than a mouse?’” says Raff. Jim Hudspeth, a neuroscientist who studies the role of hair cells in hearing at the Rockefeller University in ►

► New York, might launch a general presentation by noting the links between deafness and issues such as delayed speech in children and susceptibility to depression in adults. Then he would describe what motivates him, such as the fact that 30 million people in the United States have functionally significant hearing problems, and explain his research question: he wants to examine how structures in the ear amplify sounds.

"It's almost like a playwright coming out and saying, 'Here's a distillation of the plot, and I'm going to introduce all my actors and tell you why,'" says Christopher Nicchitta, a cell biologist at Duke University in Durham, North Carolina. "You need to build up some tension and excitement about what's going to unfold."

The actual presentation of findings can be fairly straightforward for any audience. The speaker might say simply, 'Here's what we found', and illustrate the point with some attention-grabbing data. Hudspeth might announce, for example, that he has learned that bundles of the ear's hairs respond to mechanical stimulation by exerting forces that accentuate the stimuli.

But explaining those findings — why they are significant and what they mean for the field — takes a bit more effort when the audience is less expert. Trying to build interest in discoveries is pointless if listeners ultimately say, 'So what?'

"Say you found a star that is heavier than it's supposed to be. You'd go through the conventional science wisdom about the mass of stars, how they're formed and why they should be only within a certain mass range," says De Robertis. After explaining the measuring techniques used, the speaker would show how their findings defy the widely held view about a star's maximum mass. "Then you have to explain why that's important to a larger problem — what the broader implications are," says De Robertis, adding that he might say that the estimate of the star's mass challenges what

astronomers have believed are the upper limits of mass set by outward radiation pressure.

Neither the talk's introduction, with its emphasis on context, significance and motivation, nor the explanation of implications and importance is necessary in a short, targeted talk for an audience of specialists. "You don't give the introduction — they're quite expert and they've read the abstract," says Hudspeth. "You simply say, 'Here's the scientific conundrum, here's how I'm addressing it and here's what I've found.'"

POINT AND POWERPOINT

Slides can help researchers to explain their findings. But inexperienced presenters often run into trouble by overusing colour or animation, or by cramming in too much text or data. "Don't overwhelm the audience with details," cautions Doug Cyr, a cell biologist at the University of North Carolina at Chapel Hill. Complex equations and formulae should be avoided — even if they support the discovery, says De Robertis. "Show the punchline," he says. "Leave the rest of it to the publication." Too many slides may turn an audience off. "You're not giving a commercial for your data," says Leslie Kean, a paediatric blood-and-marrow-transplant physician at Emory University in Atlanta, Georgia (see 'Power pointers').

Slides can, however, provide effective transitions between subtopics — important in longer presentations. "If you're making three or four major points, the really critical times that you can risk losing your audience are at the seams between them," says Hudspeth. He might signal to his audience that he has finished talking about anatomy, for example, by briefly summarizing what he just said. Then he might announce that a physiological analysis is next, and show a slide with a key point or a diagram that sets up the next idea. "You have to help the audience shift gears," he says.

How can speakers tell whether they are reaching their listeners? Simple — audience

audience as possible.

Be very careful about making jokes. International audiences, especially, may not understand them, and everyone will remember the joke that fell flat.

Don't swig continuously from a water bottle or jingle your change. It is distracting.

Don't get defensive or raise your voice if an audience member challenges you. Stay calm and say that it is a good point, or invite them to chat after the session.

Don't panic if you don't know the answer to a question. Reply that it is unknown, you are not sure or you had not thought of it. **K.K.**

WHAT TO AVOID

Presentation peeves

● Don't go over your allotted speaking time. The audience will be annoyed and you will cut into question time.

● Don't wear ripped jeans or shorts and sandals, even if you are in a tropical climate and living on a graduate-student budget. But don't wear a three-piece suit either, unless you are presenting to potential investors. Aim for 'business casual'.

● Don't talk quickly to fit everything in.

● Don't get very technical, even for an expert audience. Not everyone will understand.

● Only glance at the screen and at your notes. Make as much eye contact with the

members are smiling, nodding or both. By contrast, if they are tapping on phones, laptops or tablets, they are far away. If this happens, experienced presenters advise recalibrating the talk. If it seems too sophisticated, the speaker should skip some slides and change to less technical language. If the audience is more expert than anticipated, details can be added, especially if there is a whiteboard handy. Expert speakers also recommend establishing and keeping a connection by getting as physically close to the audience as possible. Hudspeth walks out from behind the lectern and comes down to the front row. Dan Agan, president of Panthera, a communications company in Alexandria, Virginia, suggests that speakers walk through the audience if they can. "You need to get closer to connect."

REACHING OUT

Presenters should also engage the audience by stimulating as many senses as possible — from sight to touch. Agan says that establishing a physical connection is key: "Do it in absentia — bring a prop, something you can pass around that's yours, and it creates the illusion of touch." Hudspeth uses a tuning fork and a 60-centimetre model of a hair bundle that shows audiences how hair cells in the ear move in response to sound waves and conduct signals to the brain. "They vary the tempo and break up the monotony of people sitting in the dark, staring at the screen," he says, "and they help make my point, especially to a general audience." If the crowd is too large to pass objects around easily, the speaker can use them as visual props and invite audience members to come up and handle them after the talk.

Speakers should also consider audience comfort: the temperature of the room; the availability of drinking water; how close together the chairs are; and the time of day. "If an audience is uncomfortable, I'll guarantee that any evaluation of my talk will go far, far down," says Agan, who adds that it is best

DOS AND DON'TS

Power pointers

- Aim to spend about one minute on each slide.
- Don't crowd in tonnes of text — it will distract the audience.
- Don't read the slide aloud — discuss a point or two.
- Use a blank, white background and an easy-to-read, unstylized typeface.
- Use animation only sparingly.
- Don't use lots of colour in your figures — it will distract from the data.
- Don't show your conclusion too early in the presentation — you will scoop yourself. **K.K.**

to present first thing in the morning, or at least before lunch. If a speaker must present immediately after lunch, and has any say in what is served, Agan recommends a light meal. Heavy fare will put an audience to sleep, he warns.

Inexperienced presenters can risk losing their audience because of unpolished technique (see 'Presentation peeves'). Rehearsing often in front of as many different groups as possible — lab mates, other postdocs or students, mentors, advisers — can mitigate the problem. Ideally, rehearsals should be filmed so that speakers can see themselves, and listeners need to be frank about recurring glitches such as repeated use of 'you know', talking fast, blinking frequently, looking down or frowning. "Ask them to be hard on you," says Nicchitta. "The more you're aware of what you're doing, the easier it is to control it so that it doesn't become a distraction."

Panic can trigger nervous mannerisms, but speakers can stave it off with a back-up plan or two. Divya Koura, a specialist in internal medicine who is doing a fellowship in oncology and haematology at Emory, gave one of her first talks in December, to a medical society. She gained confidence not only through practising for weeks in front of different groups, but also by creating a brief script. "By the end of all my rehearsals, I knew I didn't need it — it was just there," she says. "But at least I knew I was saying everything I had wanted to. There was less stumbling and no 'ums' or blank spaces."

Many speakers recommend using the 'Presenter View' feature of PowerPoint, or presenter notes in Keynote, to provide digital crib notes — safer in some cases than paper. Agan remembers watching a speaker drop a sheaf of notes in the middle of his talk. "By the time he had retrieved everything, he was so desperate and so flummoxed that his presentation turned into an indecipherable and impenetrable disaster," recalls Agan.

Seasoned presenters warn against writing out the entire talk, no matter how short — or long. It is all too easy to start reading from notes. The audience will know that they are being read to, and will drift — or, worse, leave. It is much better to create a brief outline with key points, and to rehearse the talk incessantly. Relying solely on slides can be dicey, as Raff's experience shows.

But even the best talk can suffer if the speaker doesn't use the simplest, most effective tool for establishing rapport with the audience. "The human face has 250,000 different expressions, and one stands head and shoulders above all else in terms of influencing an audience," says Agan. "And that is a smile." ■

Karen Kaplan is assistant Careers editor at Nature.

COLUMN

Ticket to everywhere

The fossilization of the PhD harms students, employers and science in general, argues **Peter Fiske**.

Many career paths are undergoing a startling transformation. Instead of locking themselves into one discipline for an entire career, professionals now switch jobs often and jump disciplines and industries routinely. Such dynamism increases productivity, creativity and entrepreneurship; interdisciplinary mobility promotes exchange of ideas and business models.

And yet the training of scientific professionals through the PhD has remained essentially unchanged since the nineteenth century. Promising students are apprenticed in academic laboratories. They learn how to conduct research, but are not prepared for success outside academia. Despite decades of policy papers, earnest admonitions and hand-wringing in the United States and Europe, actual reforms have been marginal.

The 'fossilization' of science PhD training is detrimental to young researchers — and to science as a whole. Poorly equipped to compete for jobs outside academia, many PhD holders leave university lacking confidence in their prospects. Some feel that they have bought a ticket to nowhere.

Academia might do well to look to the private sector for a model that broadens the soft skills of PhD holders and expands their prospects. Many businesses offer their executives short, intensive training programmes that stimulate their professional development in key areas such as leadership, innovation and management. Some companies develop internal schemes; others hire consultants or send their executives on week-long programmes at business schools (sometimes referred to as 'charm schools'). The goal is simple: to develop the capabilities of junior managers without costing a lot in terms of time, money or disruption to their jobs. These programmes also reflect an acknowledgement that a supervisor should not be the sole source of professional advice and mentoring.

Some research universities in the United States and Europe provide professional-development offerings for graduate students, often led by a campus career centre — a few of which have specialists dedicated to PhD students. But support is meagre: a university may have only one person to support professional development for hundreds or even thousands of PhD students. Science departments (and funding agencies) rarely provide financial support for such activities.



—ALBACHARAA-/SHUTTERSTOCK

Where such programmes exist at all, they are more tolerated than encouraged by the faculty. Academic culture enshrines the adviser–advisee relationship as the core of the PhD. But although the adviser is ideally suited to guiding students through the rigorous training necessary to become an independent researcher, he or she may be ill-equipped to help them to develop the skills to succeed in other fields. Not only do few faculty members have experience in industry, but most already have enormous demands on their time. Urging them to provide all manner of professional development is unfair to both adviser and advisee.

Formal professional-development programmes for science PhD holders should be expanded. But how to finance them? Those who benefit should pay. That includes not only funding agencies, but also students. They might pay through general student fees. Even better — to ensure that they are fully invested — they might devote credits to an actual course on career planning (see *Nature* 489, 593; 2012).

The students themselves will benefit from realizing broader career options. Universities benefit both through greater demand for their graduates and by ending up with more-satisfied (and better-paid) alumni.

But the largest benefit may be to national economies, when scientifically educated individuals enter every professional discipline — no longer because they couldn't find a job in academia, but because they chose to apply their training to important problems in a wide range of fields. With the combination of the right professional coaching and the right experience, a science PhD could turn out to be a ticket to everywhere. ■

Peter Fiske is chief executive of PAX Water Technologies in Richmond, California, and author of *Put Your Science to Work* (American Geophysical Union, 2001).

A GIFT OF PAIN

In search of a cure.

BY V. G. CAMPEN

A woman lies on the stretcher, her body rigid with pain. She watches as I manoeuvre my portable console into the curtained cubicle, keeping her eyes open until the first electrode touches her forehead.

"It's my belly that hurts," she says. "Not my head."

I respond with standardized explanations, verbiage recommended by the Association of Neural Audit Technicians. "The nerves in your body transmit sensory signals, but your brain interprets them. By monitoring your sensations I can help to make a diagnosis." She turns her head away in a dismissive movement.

"I've been trained to recognize hundreds of sensory patterns," I say, careful to avoid the word 'pain'. I remain professionally detached, neither validating nor questioning the patient, trying to minimize the confounding effect of my presence. Simply applying electrodes can soothe or inflame, depending on the patient's history and emotional state.

"It really does hurt," she whispers. "A lot."

"I document only the quality of sensations, not the intensity," I say. I'm distracted, busy stretching a baby-blue cap over my head, repositioning it until I hear the click of the magnets woven into the fabric as they mate with the metal implants in my skull. A thick umbilicus of braided wire hangs from the back of the cap. I breathe deeply and prepare for the fog — technician's slang for the disorientation that occurs when synchronizing with a patient; a wave of confusion and nausea, like waking up alone and ill in a foreign country.

I plug into the console and as the fog clears, I quite literally feel my patient's pain. Knowing I can escape by disconnecting the wires makes it bearable. I sift through the input from individual electrodes, examine various combinations, and ascribe her pain to appendicitis. It's subcostal, an atypical location, which explains why it wasn't picked up during the physical examination. And although ethically I can't report on the

intensity, privately I rate the pain an eight out of ten.

"I swear it's real," she says. "I didn't



want to come to the hospital, but it hurts so bad..."

"I believe you," I say, breaking all the rules, and tears spill onto her cheeks. "I'll find your doctor and personally give my report." I enter a brief note into her medical e-record and the system adds its automatic time stamp: midnight.

"Hey Jan, what are you doing this weekend?" Tanner, my friend and the other NAT on duty, greets me in the break room. "We're having a picnic at the lake on Sunday. Why don't you drive up with me and Keisha?" Tanner knows that Sunday is the tenth anniversary of my daughter's death.

"I'm taking a weekend shift," I say, and he nods, understanding in full.

One of the new interns, the tall skinny one, leans in through the break-room door, his arms spread to either side of the door-frame. With his baggy white coat splayed open he reminds me of a crane about to take flight.

"Um, hi," he says, looking from Tanner to me. "There's a two-year-old in bay nine, a car accident..."

"I'll take the kid," Tanner says. He puts his hand on my shoulder and walks out with

the distinctive rolling gait caused by the prosthetic blades below his knees. For some reason, people who suffer phantom limb pain tend to make excellent NATs, and the roadside bomb that tore Tanner apart and killed the others in his patrol prepared him well.

Usually I enjoy working with kids, the purity of their pain is refreshing. And yes, I know that sounds horrible. It's not something I'd say to the intern, who still has an outsider's prurient interest in NATs, undecided whether we're sadists who enjoy knowing another is in pain or masochists who enjoy feeling it. But I'm confident that each fresh-faced intern understands nociceptors and neural pathways, knows the proper use of analgesics and will learn to recognize the many forms of pain.

My therapist has asked gently, repeatedly, if my career is a way of punishing myself. I try to explain that my work isn't punishment, it's a relief. And it pays, unlike alcohol or narcotic abuse, and I've sampled both.

Here is the truth of it. My truth, at least. Physical pain pushes out the unnameable pain. The pain that has not been mapped to any region of the brain but lives in every breath. You don't need to know how my daughter died. All you need to know is that I was drinking, and she died.

I stop at the lift where a trolley bearing my midnight patient waits. She lifts her hand in greeting.

"It's appendicitis," she says. "They're admitting me for surgery." Her body is no longer contorted, probably due to painkillers, possibly aided by the comfort of a diagnosis. She grasps my hand as the lift door opens. "Thank you."

In ancient times, men known as sin-eaters would, for a few coins, eat a meal off a dying man's chest in a ritual consumption of his sins. Perhaps we are pain-eaters, no more able to conquer pain than a sin-eater could wash away sins, but giving a sliver of comfort nonetheless.

"Thank you," I say as the trolley rolls forward, pulling my patient's hand from my grasp. ■

V. G. Campen, after decades of reading, recently began writing fiction. During the day she works as a patent attorney in the biotech field.

➤ **NATURE.COM**
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

Is *Strudiella* a Devonian insect?

ARISING FROM R. Garrouste *et al.* *Nature* **488**, 82–85 (2012)

The origin of winged insects (Pterygota), one of the planet's most diverse lineages of organisms, is assumed to lie in the Devonian, but as an extremely sparse fossil record impedes our understanding of their early diversification, any well-preserved insect fossil from this time would be particularly valuable¹. Garrouste *et al.*² described an arthropod from Upper Devonian freshwater sediments from the Strud locality in Belgium (~360 million years ago) as an insect, *Strudiella devonica*. However, based on a thorough re-investigation of the specimen, we conclude that this interpretation is untenable. Hence we believe that, like several other Devonian arthropod remains^{3–7}, *Strudiella* is not an insect, and utmost diligence should be taken in interpreting poorly preserved Devonian arthropods.

Garrouste *et al.*² interpreted *Strudiella* as a pterygote nymph and reconstructed it in a suggestive illustration with a head bearing triangular (and, by implication, dicondylous) mandibles, large compound eyes, antennae with enlarged basal articles (scape, pedicel) and a slender flagellum, a thorax with three pairs of legs and an abdomen of 10–11 segments without legs. However, we are unable to corroborate any of these unambiguous insect characteristics. We cannot confirm the presence of a mandible or of mandibular teeth. The alleged teeth simply seem to be an assemblage of linearly arranged organic matter caused by the idiosyncratic way the rock parted. Indeed, the supposed teeth vanish when both part and counterpart of the fossil are considered in combination (Fig. 1b). Again, there is no evidence for the location and extent of mandibles (or for their condyles). As strongly sclerotized elements they should be easily visible even in this poorly preserved fossil. Eyes are pigment-rich organs, which are generally preserved as prominent dark spots in compression fossils. Yet, the areas sharply delineated by Garrouste *et al.*² as compound eyes are very vague and their indication of eye rims is arbitrary. Furthermore, these areas show a lower density of organic material than the surrounding areas. Therefore, in our opinion there is no evidence for large compound eyes. All in all, the entire anterior part of the fossil seems to have been just one compact unity. The reported antennae do not show an insertion on the proposed head, and we cannot confirm the subdivision into scape, pedicel and flagellum as shown in the reconstruction by Garrouste *et al.*² and which is characteristic of insects. Furthermore, the antennae are thicker than the locomotory legs, a condition highly unusual for insects (but common, for example, among crustaceans). The alleged subdivision of the remaining body into thorax and abdomen is also questionable and obscured by an uncertain number of indistinct appendages. Indeed, although the legs indicated in the original description are present, we observed additional ones beyond the three pairs usually characterizing an insect thorax (Fig. 1c, d). The abdomen shows traces of segmentation, although it is impossible to determine the number of segments (the text and illustration of Garrouste *et al.*² are in conflict as in the former 10 segments are described, whereas the latter shows 11).

In summary, we believe that Garrouste *et al.*² over-interpreted many parts of the fossil, while missing essential structures, namely additional legs. Based on our observations, its interpretation as an insect is unwarranted. The poor state of preservation allows for a number of alternative interpretations, such as a decayed crustacean. *Strudiella* is a very unusual fossil for the Strud locality and its interpretation remains a challenging task until better preserved specimens

are discovered. Hence, we consider it to be crucial to prevent this fossil from entering entomology textbooks as an early insect.

METHODS

Beyond direct observation (stereomicroscope Zeiss Stemi SV11 Apo with Plan Apochromat S ×1.6 and camera lucida) we applied high-resolution photography (compound microscope Leica CM E with ×4 and ×10 lenses and 5 megapixel digital camera Scopetek DCM 510) in combination with polarized light as well as stitching and focus stacking techniques (Helicon Focus 4.2.9 ×64, Photoshop CS3). The re-investigation was made possible by G. Clément. We did not conduct any preparation of the specimen.

Thomas Hörnschemeyer¹, Joachim T. Haug², Olivier Bethoux³, Rolf G. Beutel⁴, Sylvain Charbonnier³, Thomas A. Hegna⁵, Markus Koch⁶, Jes Rust⁷, Sonja Wedmann⁸, Sven Bradler¹ & Rainer Willmann¹

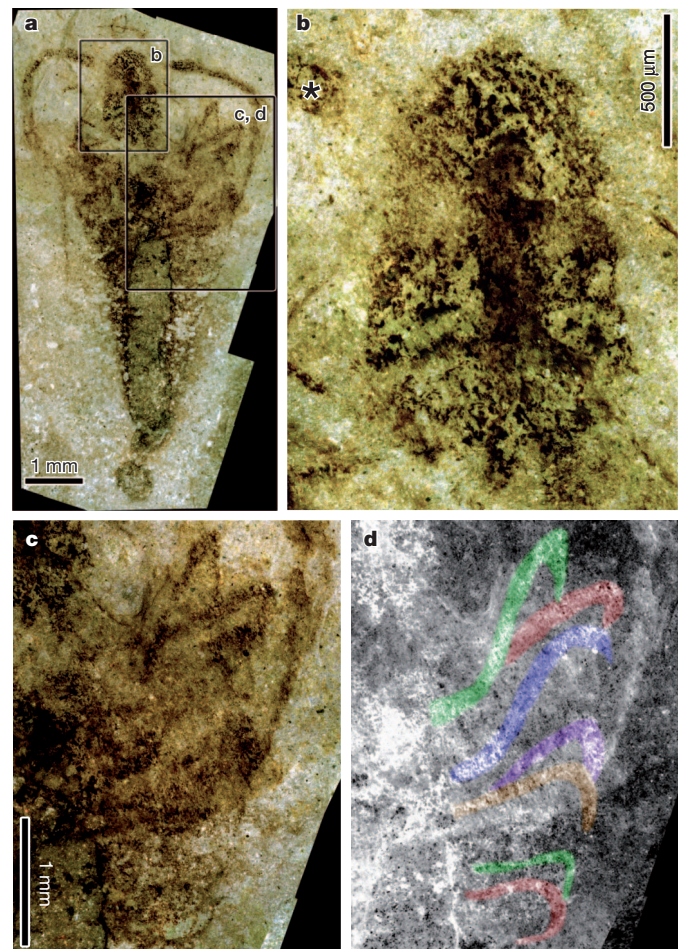


Figure 1 | *Strudiella devonica* Garrouste *et al.* 2012. **a**, Photograph of the holotype (Institut Royal des Sciences Naturelles de Belgique, a12818) based on part and counterpart (stitching and focus stacking procedure). **b**, Detail of the proposed head based on part and counterpart. *, base of hypothetical antenna. **c**, Detail of legs. **d**, Same area as in **c**, inverted grey-scale with some legs colour-coded.

¹Johann-Friedrich-Blumenbach-Institute of Zoology and Anthropology, Georg-August-University Göttingen, 37073 Göttingen, Germany. email: rwillma1@gwdg.de

²Zoological Institute and Museum, Department of Cytology and Evolutionary Biology, University of Greifswald, 17487 Greifswald, Germany.

³Muséum National d'Histoire Naturelle UMR7207 CNRS (CR2P) CP38, 75231 Paris Cedex 05, France.

⁴Entomology Group, Institut für Spezielle Zoologie und Evolutionsbiologie, Friedrich-Schiller-University Jena, 07743 Jena, Germany.

⁵Geology Department, Western Illinois University, 1 University Circle, 113 Tillman Hall, Macomb, Illinois 61455, USA.

⁶Biocentre Grindel and Zoological Museum, University of Hamburg, 20146 Hamburg, Germany.

⁷Steinmann Institute, Department of Palaeontology, University of Bonn, 53115 Bonn, Germany.

⁸Senckenberg Forschungsinstitut und Naturmuseum, Forschungsstation Grube Messel, 64409 Messel, Germany.

Received 30 October; accepted 5 December 2012.

1. Shear, W. A. An insect to fill the gap. *Nature* **488**, 34–35 (2012).
2. Garrouste, R. *et al.* A complete insect from the Late Devonian period. *Nature* **488**, 82–85 (2012).
3. Willmann, R. Reinterpretation of an alleged marine hexapod stem-group representative. *Org. Divers. Evol.* **5**, 199–202 (2005).
4. Köhl, G. & Rust, J. *Devonohexapodus bocksbergensis* is a synonym of *Wingertshellicus backesi* (Euarthropoda) – no evidence for marine hexapods living in the Devonian Hunsrück Sea. *Org. Divers. Evol.* **9**, 215–231 (2009).
5. Hennig, W. *Insect Phylogeny* (Wiley, 1981).
6. Labandeira, C. C., Beall, B. S. & Hueber, F. M. Early insect diversification: Evidence from a Lower Devonian bristletail from Québec. *Science* **242**, 913–916 (1988).
7. Jeram, A. J., Selden, P. A. & Edwards, D. Land animals in the Silurian: arachnids and myriapods from Shropshire, England. *Science* **250**, 658–661 (1990).

Author Contributions T.H., J.T.H., O.B., S.B. and R.W. examined the fossil, T.H., J.T.H., O.B., T.A.H., S.W., S.B. and R.W. drafted the manuscript, T.H., O.B., R.G.B., S.C., T.A.H., M.K., J.R., S.B. and R.W. edited the final article.

Competing Financial Interests Declared none.

doi:10.1038/nature11887

Garrouste *et al.* reply

REPLYING TO T. Hörschemeyer *et al.* *Nature* **494**, <http://dx.doi.org/10.1038/nature11887> (2013)

Since the nineteenth century, Devonian insects have repeatedly proven to be something else^{1–5}, the sole exception being *Rhyniognatha*. Recently the Devonian insect *Strudiella devonica*⁶ has been denied by Hörschemeyer *et al.*⁷, who could not “confirm the presence of a mandible or of mandibular teeth”.

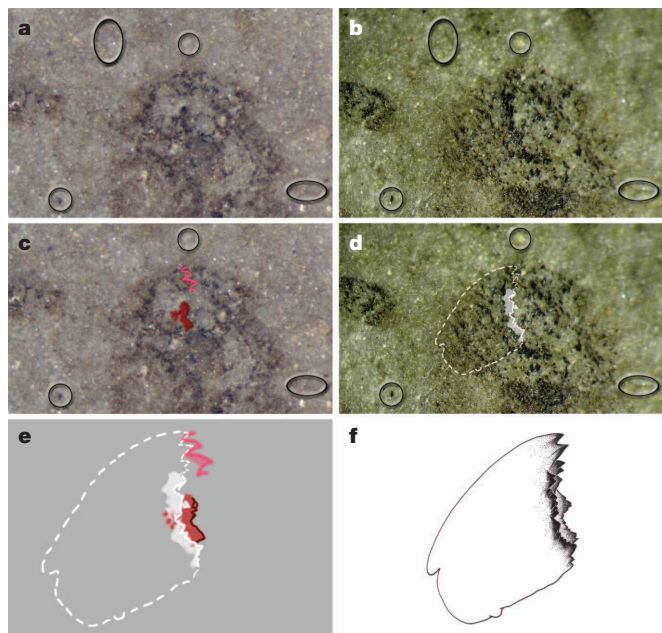


Figure 1 | *Strudiella devonica* Garrouste *et al.* 2012. **a–d**, Photographs of anterior portion of head, holotype IRSNB a12818, with reference points indicated by ellipses. **a, c**, Counterprint in symmetry, showing right incisor teeth (pink) and second rank of teeth of left molar plate (red). **b, d**, Print of same region showing left mandible, molar plate (white). **e**, Reconstruction of the mandibles showing partly overlapping incisors and typical left molar plate. **f**, Proposal of three-dimensional reconstruction of the left mandible, thanks to G. Hodebert.

However, Figure 1b by Hörschemeyer *et al.*⁷ supports the interpretation of the presence of mandibles because it reveals the incisor part of the right mandible, not massive but sharp and in the precise location where it should be, partly overlapping the left incisors (Fig. 1a, c, e). Comparison of part and counterpart with accurate reference points reveals that the left mandible has a massive molar typical of insects (Fig. 1b–e). They do not discuss the clearly visible palps, abnormal under a crustacean hypothesis. They “cannot confirm the subdivision into scape, pedicel and flagellum” of the antenna, but we believe that the identity of lengths and widths of the corresponding left and right antennomeres perfectly delimitate them. Scape and pedicel are thicker than flagellomeres. That the antennae are thicker than the legs as preserved, although unusual, is not contradictory with a hexapod attribution and would constitute an autapomorphy for *Strudiella* among Palaeozoic genera. Eye zones are darker than areas near the mandibles but with some parts removed and missing, a feature quite frequent in fossil insects in which eye facets are rarely preserved. The subdivision into three tagmata is quite clear, the ten abdominal segments being individualized by a median higher line corresponding to digestive organs. If they confirm our interpretation of thoracic legs, no word is said against the long femora and tibiae, typical of insects but not of basal hexapods and Crustacea. Their alleged supplementary legs have very different shapes, with no angles between their very short and curved ‘spaghetti-like’ structures, and would apparently not have been articulated into podites. These asymmetrical soft-tissue structures are probably internal organs extruded during compression and decay.

We believe that Hörschemeyer *et al.*⁷ fail to contradict the clear subdivision into three tagmata, the presence of insect mandibles, or of only one pair of antennae with thickened scape and pedicel, and to demonstrate that *Strudiella* has supernumerary legs. Instead, their new images reaffirm the presence of the right mandible. Because of its decayed condition and absence of wings, *Strudiella* reaffirms notions of Late Devonian insect diversity but does not contribute to solving the ‘wing enigma’.

BRIEF COMMUNICATIONS ARISING

Romain Garrouste¹, Gaël Clément², Patricia Nel¹, Michael S. Engel³, Philippe Grandcolas¹, Cyrille A. D'Haese¹, Linda Lagebro⁴, Julien Denayer⁵, Pierre Gueriau^{2,6}, Patrick Lafaite⁷, Sébastien Olive^{5,8}, Cyrille Prestianni⁸ & André Nel¹

¹UMR CNRS 7205, CP 50, Entomologie, Muséum national d'Histoire Naturelle, 45 rue Buffon, F-75005 Paris, France.

email: anel@mnhn.fr

²UMR CNRS 7207, Paléontologie, Muséum national d'Histoire Naturelle, 8 rue Buffon, F-75005 Paris, France.

³Division of Entomology, Natural History Museum, and Department of Ecology and Evolutionary Biology, 1501 Crestline Drive – Suite 140, University of Kansas, Lawrence, Kansas 66045, USA.

⁴Department of Earth Sciences, Uppsala University, Villavägen 16, SE-752 36 Uppsala, Sweden.

⁵Service de Paléontologie animale et humaine, Département de Géologie, Université de Liège, Bât. B.18, Allée du Six-Août, Sart Tilman, B-4000 Liège, Belgium.

⁶IPANEMA, USR3461CNRS- Ministère de la Culture et de la Communication, F-91190, Saint-Aubin, France.

⁷CNRS-MNHN DICAP, Service Multimedias, CP 27, 57 rue Cuvier, F-75005 Paris, France.

⁸Royal Belgian Institute of Natural Sciences Paleontology Department, 29, Rue Vautier, B-1000 Brussels, Belgium.

1. Brongniart, C. Insectes fossiles des grès siluriens. *La Nature, Paris* **13**, 116 (1885).
2. Agnus, A. *Palaeoblattina douvillei*, considéré d'abord comme un insecte est une pointe génale de trilobite. *C. R. Acad. Sci., Paris* **138**, 398 (1904).
3. Sharov, A. G. Peculiar Paleozoic insects of the new order Monura (Insecta, Apterygota) [in Russian]. *Doklady Akademii nauk. SSSR* **115**, 795–798 (1957).
4. Rohdendorf, B. B. Devonian eopterids were not insects but Eumalacostraca (Crustacea) [English translation]. *Entomol. Rev.* **51**, 96–97 (1972).
5. Kühl, G. & Rust, J. *Devonohexapodus bocksbergensis* is a synonym of *Wingertshellicus backesi* (Euarthropoda) – no evidence for marine hexapods living in the Devonian Hunsrück Sea. *Org. Divers. Evol.* **9**, 215–231 (2009).
6. Garrouste, R. *et al.* A complete insect from the Late Devonian period. *Nature* **488**, 82–85 (2012).
7. Hörnschemeyer, T. *et al.* Is *Strudiella* a Devonian insect? *Nature* **494**, <http://dx.doi.org/10.1038/nature11887> (2013).

Author Contributions R.G. and P.N. are first authors with equal rank; R.G., A.N., P.N., P.Gr., C.A.D'H., L.L., M.S.E., J.D., C.P., P.Gu. and S.O. drafted the manuscript and prepared figures. A.N. and P.N. coordinated the manuscript; G.C. coordinated and participated in fieldwork at the Strud locality and contributed to draft the manuscript; L.L., J.D., C.P., P.Gu., P.L. and S.O. also participated in fieldwork.

doi:10.1038/nature11888